

## King's Research Portal

DOI:

[10.1161/CIRCGENETICS.116.001487](https://doi.org/10.1161/CIRCGENETICS.116.001487)

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Hedman, Å. K., Mendelson, M. M., Marioni, R. E., Gustafsson, S., Joeanes, R., Irvin, M. R., Zhi, D., Sandling, J. K., Yao, C., Liu, C., Liang, L., Huan, T., McRae, A. F., Demissie, S., Shah, S., Starr, J. M., Cupples, L. A., Deloukas, P., Spector, T. D., ... Ingelsson, E. (2017). Epigenetic Patterns in Blood Associated with Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results from Genome-Wide Association Studies. *Circulation-Cardiovascular Genetics*, 10(1), [001487].  
<https://doi.org/10.1161/CIRCGENETICS.116.001487>

### Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results From Genome-Wide Association Studies

Åsa K. Hedman, PhD\*; Michael M. Mendelson, MD\*; Riccardo E. Marioni, PhD; Stefan Gustafsson, PhD; Roby Joehanes, PhD; Marguerite R. Irvin, PhD; Degui Zhi, PhD; Johanna K. Sandling, PhD; Chen Yao, PhD; Chunyu Liu, PhD; Liming Liang, PhD; Tianxiao Huan, PhD; Allan F. McRae, PhD; Serkalem Demissie, PhD; Sonia Shah, PhD; John M. Starr, MD, PhD; L. Adrienne Cupples, PhD; Panos Deloukas, PhD; Timothy D. Spector, MD; Johan Sundström, MD, PhD; Ronald M. Krauss, MD; Donna K. Arnett, PhD; Ian J. Deary, PhD; Lars Lind, MD, PhD; Daniel Levy, MD†; Erik Ingelsson, MD, PhD†

**Background**—Genome-wide association studies have identified loci influencing circulating lipid concentrations in humans; further information on novel contributing genes, pathways, and biology may be gained through studies of epigenetic modifications.

**Methods and Results**—To identify epigenetic changes associated with lipid concentrations, we assayed genome-wide DNA methylation at cytosine–guanine dinucleotides (CpGs) in whole blood from 2306 individuals from 2 population-based cohorts, with replication of findings in 2025 additional individuals. We identified 193 CpGs associated with lipid levels in the discovery stage ( $P < 1.08 \times 10^{-7}$ ) and replicated 33 (at Bonferroni-corrected  $P < 0.05$ ), including 25 novel CpGs not previously associated with lipids. Genes at lipid-associated CpGs were enriched in lipid and amino acid metabolism processes. A differentially methylated locus associated with triglycerides and high-density lipoprotein cholesterol (HDL-C; cg27243685;  $P = 8.1 \times 10^{-26}$  and  $9.3 \times 10^{-19}$ ) was associated with *cis*-expression of a reverse cholesterol transporter (*ABCG1*;  $P = 7.2 \times 10^{-28}$ ) and incident cardiovascular disease events (hazard ratio per SD increment, 1.38; 95% confidence interval, 1.15–1.66;  $P = 0.0007$ ). We found significant *cis*-methylation quantitative trait loci at 64% of the 193 CpGs with an enrichment of signals from genome-wide association studies of lipid levels ( $P_{TC} = 0.004$ ,  $P_{HDL-C} = 0.008$  and  $P_{triglycerides} = 0.00003$ ) and coronary heart disease ( $P = 0.0007$ ). For example, genome-wide significant variants associated with low-density lipoprotein cholesterol and coronary heart disease at *APOB* were *cis*-methylation quantitative trait loci for a low-density lipoprotein cholesterol–related differentially methylated locus.

**Conclusions**—We report novel associations of DNA methylation with lipid levels, describe epigenetic mechanisms related to previous genome-wide association studies discoveries, and provide evidence implicating epigenetic regulation of reverse cholesterol transport in blood in relation to occurrence of cardiovascular disease events. (*Circ Cardiovasc Genet.* 2017;10:e001487. DOI: 10.1161/CIRCGENETICS.116.001487.)

**Key Words:** cardiovascular diseases ■ DNA Methylation ■ epigenomics ■ gene expression ■ lipids

Cardiovascular disease (CVD) is the leading cause of death worldwide.<sup>1</sup> Serum concentrations of total cholesterol (TC) and subcomponents of low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides are established risk factors for coronary heart disease (CHD).<sup>1</sup> Recent studies have provided evidence of causal roles for LDL-C and triglycerides in CHD.<sup>2,3</sup> Further understanding of the genomic regulatory mechanisms linking lipids to CHD may enhance our ability to predict CHD risk, tailor current CHD treatments, or discover new treatments for CHD.

## See Clinical Perspective

Genome-wide association studies (GWAS) have been successful in identifying numerous single-nucleotide polymorphisms (SNPs) associated with lipid levels and CHD.<sup>4,5</sup> Because many of the SNPs are located in noncoding regions, epigenetic mechanisms can be suspected to mediate many of the genetic discoveries. Integrative analyses of methylation of cytosine nucleotides at cytosine–guanine dinucleotide (CpG) sites with genetic sequence variants and gene expression may elucidate previously unknown genes and pathways underlying GWAS discoveries. In

Received February 23, 2016; accepted November 14, 2016.

Guest Editor for this article was Christopher Semsarian, MBBS, PhD, MPH.

\*Drs Hedman and Mendelson contributed equally to this work first authors.

†Drs Levy and Ingelsson are joint senior authors.

The Data Supplement is available at <http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.116.001487/-/DC1>.

Correspondence to Erik Ingelsson, MD, PhD, Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, 300 Pasteur Dr, Mail code 5773, Stanford, CA 94305. E-mail [eriking@stanford.edu](mailto:eriking@stanford.edu)

© 2017 The Authors. *Circulation: Cardiovascular Genetics* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited.

*Circ Cardiovasc Genet* is available at <http://circgenetics.ahajournals.org>

DOI: 10.1161/CIRCGENETICS.116.001487

addition to variation in DNA methylation that is determined by the surrounding genetic sequence,<sup>6</sup> methylation is also affected by early exposures in utero<sup>7,8</sup> and later life environmental factors.<sup>9,10</sup> Environmentally induced alterations in DNA methylation may mediate environmental contributions to disease<sup>11</sup> and reveal novel genes and pathways involved in disease that cannot be discovered in GWAS alone. Regulation of gene expression via DNA methylation may explain an additional component of interindividual variation in lipid levels beyond genetic sequence variants. Because much of the population burden of dyslipidemia and CHD is not explained by GWAS loci, relating differential DNA methylation to gene expression, intermediate metabolites, and disease end points may be useful in identifying additional candidate genes and mechanisms for which directed perturbation may help prevent morbidity and mortality from CHD.

In this study, we aimed to identify epigenetic variation in relation to lipid levels through epigenome-wide association analyses of whole blood–derived DNA in  $\leq 2306$  individuals with independent external replication of findings in  $\leq 2025$  individuals. Methylation differences in blood-derived DNA have been shown to reflect transtissue differential methylation in various tissues,<sup>12–14</sup> including liver<sup>15</sup> and adipose.<sup>16</sup> In addition to the discovery of lipid-related differential DNA methylation, we assessed the association of lipid-related epigenetic changes to the risk of incident CHD events. Finally, we combined lipid-associated DNA methylation with genetic sequence variants, gene expression, and intermediate metabolites in an attempt to unravel the underlying genomic regulatory mechanisms linking serum lipid measures to CHD risk.

## Methods

### Study Participants and Design

We conducted an epigenome-wide association study of serum lipid concentrations (TC, HDL-C, LDL-C, and triglycerides) in over 4000 adult participants from large community-based cohorts in the United States and Europe (Figure 1). Ethical approvals for the project were granted by the local Ethics Committee for each of the participating cohorts, and all samples were collected after obtaining written and signed informed consent. Participants from the FHS (Framingham Heart Study) offspring cohort ( $n=1494$ ; mean [SD] age=66.4 [8.9] years)<sup>17</sup> and the PIVUS (Prospective Investigation of the Vasculature in Uppsala Seniors Study;  $n\leq 812$ ; 70.2 [0.2] years)<sup>18</sup> were included in the discovery analysis. Loci identified as significant in the discovery ( $P<1.08E-07$ ; Bonferroni-adjusted  $P$  value for multiple testing) were then examined for external replication in participants from the LBC1921 (Lothian Birth Cohorts of 1921;  $n\leq 380$ ; 79.1 [0.6] years) and LBC1936 (LBC of 1936;  $n\leq 654$ ; 69.5 [0.8] years)<sup>19–21</sup> and the GOLDN (Genetics of Lipid Lowering Drugs and Diet Network;  $n=991$ ; 48.8 [16] years).<sup>22</sup> Characteristics of the cohorts are available in Table I in the [Data Supplement](#). Further details about cohort-specific study design and sample collection are available in Methods in the [Data Supplement](#). Primary analyses examined the association of each lipid component with methylation levels in blood at 459433 CpGs and were adjusted for age, sex, white cell counts (if applicable), and batch effects; secondary models additionally adjusted for body mass index (BMI). We excluded individuals taking lipid medications (statins, fibrates, etc.) because the cross-sectional design would not allow us to determine if DNA methylation changes contributed to elevated lipids necessitating lipid medications or were secondary to medication use. The identified differentially methylation loci were assessed for associations with nearby genetic sequence variants in *cis* (defined as  $\pm 100$  kb), intermediate phenotypes (gene expression and metabolites in blood), and incident CHD events.

### Phenotype Measurements and Disease Outcomes

Lipids traits were measured in blood samples collected after fasting with the exception of LBC (LBC1921 and LBC1936) for which non-fasting blood was drawn. Lipid measurements were performed using standard methods as described in Methods in the [Data Supplement](#) for each study. In FHS, PIVUS, and LBC1936, LDL-C levels were calculated by the Friedewald equation, whereas levels were directly measured in GOLDN. In LBC1921, HDL-C and LDL-C were not available. Characteristics of the lipid traits for each cohort are available in Table I in the [Data Supplement](#). Weight and height were measured in each study using standardized protocols. BMI was calculated as weight in kg divided by height in m<sup>2</sup>. In FHS and PIVUS, cardiovascular events during  $\leq 10$  years of follow-up ( adjudicated by physicians) were used to define a composite CHD end point, which included fatal or nonfatal myocardial infarction and revascularization procedure (percutaneous transluminal coronary angioplasty or coronary artery bypass graft). In FHS, data on coronary death and coronary insufficiency (unstable angina) were also included.

### Genome-Wide DNA Methylation Profiling

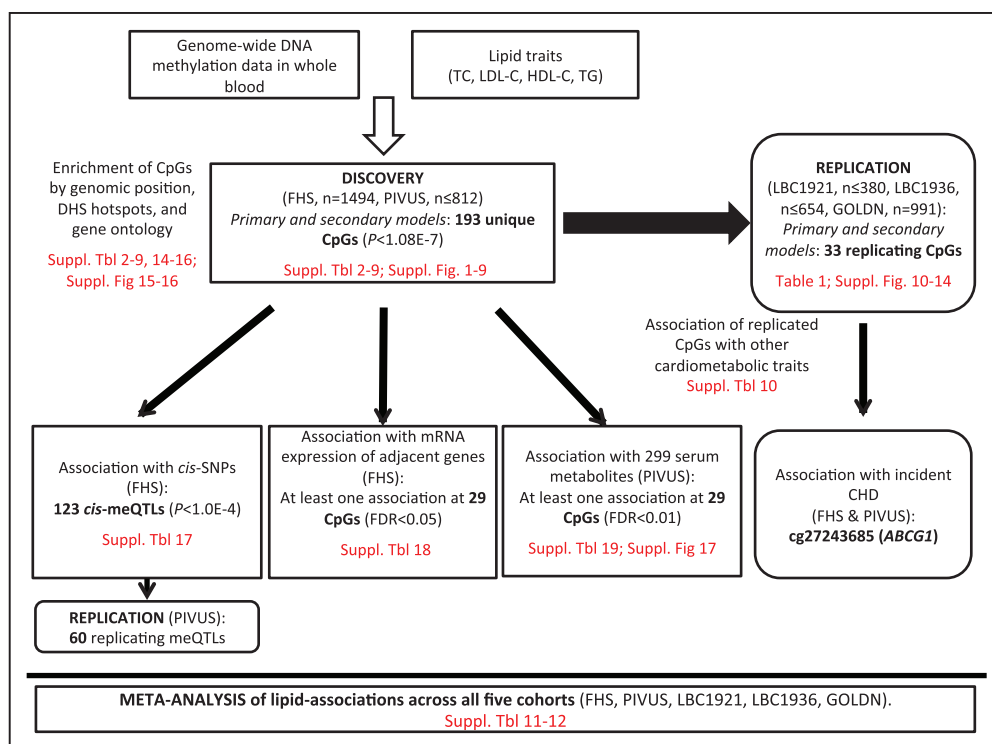
Genome-wide DNA methylation profiling was performed on genomic DNA isolated from whole blood (FHS, PIVUS, LBC1921, and LBC1936) or CD4<sup>+</sup> T cells (GOLDN). DNA samples were bisulphite converted and analyzed on Illumina HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA) following the manufactures' protocol. After quality control procedures, methylation data were available for analyses in 2377 FHS, 967 PIVUS, 446 LBC1921, 920 LBC1936, and 995 GOLDN participants. Further cohort-specific details and quality control procedures are available in Methods in the [Data Supplement](#). In all studies, blood used in extraction of DNA for methylation analysis was collected at the same time point as phenotype and covariate measurements.

### Additional Molecular Genomics Data

In FHS, SNP data were obtained from the Affymetrix 550K Array (Affymetrix, Santa Clara, CA) and imputed to 1000 Genomes SNPs (phase 1 release), as previously reported.<sup>23</sup> The FHS genotype data are available at Database of Genotypes and Phenotypes under the accession number phs000342.v13.p9. In PIVUS, individuals were genotyped using the Illumina OmniExpress and Illumina MetaboChip microarrays. Data were imputed to 1000G (version: March 2012) using Impute v.2.2.2.<sup>24</sup> Gene expression profiles in blood, obtained using the Affymetrix Human Exon 1.0 ST GeneChip platform, were available for 2246 participants in the FHS. Untargeted metabolomic profiles in serum were available for 785 PIVUS participants also included in the lipid-association analyses. Acquity Ultra Performance Liquid Chromatography coupled to a Xevo G2 Q-TOFMS (Waters Corporation, Milford, MA) was used in metabolomic profiling. Only annotated metabolites ( $n=229$ ) were used in analysis in relation to DNA methylation. Further details are available in Methods in the [Data Supplement](#).

### Annotation of DNA Methylation Probes

Mapping and annotation of the 485764 probes on the HumanMethylation450K BeadChip have previously been described.<sup>25</sup> Only autosomal probes were included in analyses. Briefly, probes mapping to multiple locations (with at least 2 mismatches) in the human reference genome (GRCh37) were excluded. Furthermore, probes were filtered based on SNPs as follows: those with a common SNP (minor allele frequency  $>5\%$ ) within 10 bp of the methylation site and those overlapping copy number variants were excluded from analysis. This resulted in a final set of probes which were assigned to CpG islands and RefSeq transcripts downloaded from the UCSC Genome Browser. Probes within 2 kb away from borders of a CpG island were defined as shores and those within 2 kb of shores as falling within shelves. The rest were assigned to others/open sea. Probes were mapped in relation to transcripts as follows: TSS1500 (1500–200 bp upstream of transcription start site), TSS200 (200 bp upstream of transcription start site), the 5'-UTR (untranslated region), the first exon, the gene body, or the 3'-UTR.<sup>26</sup>



**Figure 1.** Overview of the study. CpG indicates cytosine–guanine dinucleotide; FDR, false discovery rate; FHS, Framingham Heart Study; GOLDN, Genetics of Lipid Lowering Drugs and Diet Network; HDL-C, high-density lipoprotein cholesterol; LBC, Lothian Birth Cohorts; LDL-C, low-density lipoprotein cholesterol; meQTL, methylation quantitative trait locus; PIVUS, Prospective Investigation of the Vascu-  
lature in Uppsala Seniors Study; TC, total cholesterol; and TG, triglyceride.

## Statistical Analysis

### Association of Methylation of Blood Cell–Derived DNA With Lipids

Multivariable linear regression models were conducted (using cohort-specific approaches described in Methods in the [Data Supplement](#)) with DNA methylation  $\beta$  value specified as the dependent variable and the lipid component as the independent variable of interest. The primary model was adjusted for age, sex, white cell count (if applicable), technical covariates, and, if applicable, family structure (included as random effects using the R packages *pedigreemm*<sup>27</sup> [FHS] or *kinship*<sup>28</sup> [GOLDN], see further details in Methods in the [Data Supplement](#)). Secondary models additionally adjusted for BMI. Individuals on lipid-lowering medications were excluded from all analyses. Lipid levels (in mg/dL) were analyzed on the raw scale, except levels of triglyceride that were natural log-transformed before analyses. Probes with a common SNP (minor allele frequency >5%) within 10 bp of the methylation site were excluded from analysis. Fixed-effect meta-analyses were performed using the inverse variance–weighted method implemented in METAL<sup>29</sup> of genome-wide association results in the discovery cohorts (FHS and PIVUS). CpGs significant at Bonferroni-corrected  $\alpha$  threshold <0.05 (taking the number of CpGs into account; corresponding to a nominal  $P < 1.08E-7$ ) in discovery were analyzed in the replication cohorts. Meta-analyses of the results in the individual replication cohorts (LBC1921, LBC1936, and GOLDN) were performed using the same method as above.

### Cross-Tissue Validation of Lipid Associations

Lipid-associated CpGs in blood were validated in DNA methylation data from subcutaneous abdominal adipose tissue (SAT) from the MuTHER (Multiple Tissue Human Expression Resource) study.<sup>30</sup> This study and data set is described in detail in Grundberg et al.<sup>25</sup> The study contains genome-wide DNA methylation data using the Illumina HumanMethylation450 array collected from 648 female twins and singletons (97 monozygotic pairs, 162 dizygotic pairs, and 130 singletons) of European ancestry. The participants had a mean

age of  $\approx 60$  years and a mean BMI of 26.6 kg/m<sup>2</sup>. After removing individuals on lipid-lowering medication and with missing phenotype, a total of 588, 588, 589, and 639 participants were considered in the analyses of TC, LDL-C, HDL-C, and triglycerides, respectively. For association with phenotype, a linear mixed effects model was fitted which was adjusted for age, bisulphite conversion concentration, bisulphite conversion efficiency, and BeadChip as fixed effects and family relationship (twin pairing) and zygosity as random effects. One-hundred sixty-four out of 193 lipid-associated CpGs could be tested in SAT.

### Gene Set Enrichment Analysis

To place our data in the context of biological processes or pathways, we subjected genes annotated to CpG sites (from 1500 bp upstream of transcription start site to 3'-UTR)<sup>26</sup> associated with phenotypes to pathway analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID).<sup>31,32</sup> We used annotations from the Kyoto Encyclopedia of Genes and Genomes, Protein Analysis Through Evolutionary Relationships, Gene Ontology, REACTOME, and Clusters of Orthologous Groups of proteins.

### Methylation Quantitative Trait Locus Analysis

Methylation quantitative trait locus (meQTL) analysis for lipid-associated methylation probes was performed in the FHS cohort (n=2246), and significant lead meQTL SNPs ( $P < 1E-04$ ) were tested for replication in the PIVUS cohort (n=775). MeQTL analysis was limited to SNPs located within 100 kb either side of the probe location (*cis*) and SNPs with a minor allele frequency >5% and imputation quality  $R_{sq} > 0.8$ . In FHS, the residual of the DNA methylation  $\beta$  value was extracted after the removal of the fixed (age, sex, and imputed white cell counts using the Houseman method<sup>33</sup>) and random covariates (chip, row, and column), along with the kinship correlation structure. The DNA methylation residual was regressed on the SNP genotype additionally adjusting for 25 methylation principal components to account for unmeasured technical variation. Imputed SNPs were entered into the model as allele dosages. In PIVUS, the



association between normalized methylation  $\beta$  values and posterior mean genotypes (MACH format) was modeled by a linear mixed effect model, using R<sup>34</sup> and the *lmer* function (lme4 package), fitted by maximum likelihood assuming a normally distributed error term. Models were adjusted for age, sex, and predicted white cell counts (estimated from the DNA methylation data using the Houseman algorithm<sup>33</sup> as implemented in R package *minfi*<sup>35</sup>) as fixed effects and chip, chip row, and chip column as random effects.

#### Association With Gene Expression Data

In FHS, the association between DNA methylation and gene expression (available in 2246 participants with DNA methylation) was performed on the gene expression residuals after the removal of the fixed and random covariates, along with the kinship correlation structure using a linear model, primarily to avoid potential confounding by blood count. Only CpGs that were methylome-wide significant were tested, and individual CpGs were tested against a single gene expression transcript in the regression model. All gene transcripts within  $\pm 500$  kb (cis) of the CpG were assessed.

#### Association With Targeted Metabolites

In PIVUS, the associations between normalized methylation  $\beta$  values at lipid-associated CpGs and 229 serum metabolites were modeled by a linear mixed effect model, using R<sup>34</sup> and the *lmer* function (lme4 package), fitted by maximum likelihood assuming a normally distributed error term. Models were adjusted for age, sex, and predicted white cell counts (using the Houseman algorithm<sup>33</sup> in R package *minfi*<sup>35</sup>) as fixed effects and chip, chip row, and chip column as random effects. False discovery rate (FDR) were estimated based on  $Q$  values.<sup>36</sup>

#### Association With Disease Outcome

In FHS, Cox models were fitted in R using the *coxme* package to model the association of baseline DNA methylation with incident CHD events adjusted for age, sex (fixed effects), and family structure (mixed effect) for the 33 replicated lipid-associated CpGs. As using measured technical covariates (chip, row, and column) with a binary outcome resulted in too many overall levels, surrogate variable analysis (that capture sources of heterogeneity in the methylation data and can be used to control for the influence of these latent variables on inference)<sup>37</sup> was used to capture the measured and unmeasured technical variation in the methylation data, and 5 surrogate variables (associated with incident CHD at  $P$  value  $< 0.05$ ) were included as covariates in the model.

In PIVUS, Cox models were fitted in R using the *coxph* function in the survival package, to model the association between case/control status and standardized methylation levels at the 33 replicated lipid-associated CpGs. Models were adjusted for age, sex, chip, and predicted white cell counts (using the Houseman algorithm<sup>33</sup> in the R package *minfi*<sup>35</sup>).

## Results

### Associations of DNA Methylation With Lipid Levels in Blood

We sought to examine whether differences in DNA methylation were associated with circulating lipid levels (study design and main results outlined in Figure 1). After meta-analysis of 459 433 CpGs in the FHS ( $n=1494$ ) and PIVUS ( $n=812$ ) studies, we found methylation at 40, 23, 110, and 28 CpG sites associated with TC, LDL-C, HDL-C, and triglycerides, respectively, at methylome-wide significant threshold ( $P < 1.08 \times 10^{-7}$ ; Volcano plots in Figures I through IV in the [Data Supplement](#); Manhattan plots in Figures V through VIII in the [Data Supplement](#)). In total, there were 184 unique CpG sites (annotated to 138 unique genes) associated with any lipid level (some were associated with several); 174 of these have not previously been reported to be associated with lipid levels. Complete results are available in Tables II through V in

the [Data Supplement](#), and the level of overlap between CpGs associated with the 4 lipid fractions is depicted in Figure IXa in the [Data Supplement](#).

In secondary analyses additionally adjusted for BMI, 80% (32/40), 87% (20/23), 13% (14/110), and 61% (17/28) of the CpG sites associated in the primary model with TC, LDL-C, HDL-C, and triglycerides, respectively, were significantly associated in the corresponding BMI-adjusted lipid model at a methylome-wide significant threshold ( $P < 1.08 \times 10^{-7}$ ; Volcano plots in Figures I through IV in the [Data Supplement](#)). Associations of methylation with lipid levels after adjustment for BMI occurred at 80 unique CpGs (annotated to 60 unique genes). In these BMI-adjusted analyses, we found 9 CpG sites associated with lipid levels that were not significantly associated in the primary analyses (complete results available in Tables VI through IX in the [Data Supplement](#); Figure IXb in the [Data Supplement](#)).

We then attempted to replicate the associations at the 193 CpG sites significantly associated with at least 1 lipid trait (in models without or with BMI adjustment) in 3 independent cohorts ( $\leq 2025$  individuals) with DNA methylation from whole blood (LBC1936 and LBC1921) or CD4<sup>+</sup> T cells (GOLDN). At a Bonferroni-corrected  $\alpha$  threshold of 0.05 (taking the number of tests per lipid trait into account) and taking direction of effect into account, 5 (13%), 1 (4%), 11 (10%), and 19 (68%) of the CpG sites associated with TC, LDL-C, HDL-C and triglycerides, respectively, in the primary analysis replicated in a meta-analysis of these 3 independent cohorts (Table 1). When only considering the 10 most associated CpGs in the discovery for each lipid trait, the replication rate was considerably higher (30%, 10%, 40%, and 90% for TC, LDL-C, HDL-C, and triglycerides, respectively). Comparison of effect sizes between discovery and replication for all CpGs significant in the discovery stage revealed a high degree of overall concordance between the  $\beta$  coefficients (Pearson correlation coefficients 0.78, 0.67, 0.71, and 0.88, for TC, LDL-C, HDL-C, and triglycerides, respectively), indicating a high level of agreement even for CpGs that did not replicate at the  $P$  value threshold (Figure X in the [Data Supplement](#)). Comparison of effect sizes between discovery and each of the individual replication cohorts for all CpGs significant in the discovery is included in Figures XI through XIV in the [Data Supplement](#). In secondary analyses adjusted for BMI in the external cohorts, we replicated 4 (13%), 1 (5%), 2 (14%), and 12 (71%) of the CpG sites associated with TC, LDL-C, HDL-C, and triglycerides, respectively (Table 1). In total, 33 CpGs replicated in the primary or secondary model (representing 55 associations as some CpGs were associated with several lipid traits). Twenty-five of these have not previously been reported to be associated with lipids in DNA methylation studies (Table 1; Table X in the [Data Supplement](#)). Ten of the lipid-associated CpGs (including 5 of the novel CpGs) have previously been associated with adiposity (BMI and waist circumference), glycemic traits (fasting insulin and insulin resistance by homeostasis model assessment), or type 2 diabetes mellitus in blood cell-derived DNA methylation data (Table X in the [Data Supplement](#)). We tested whether associations in blood could also be detected in another tissue using DNA methylation data from abdominal SAT from the MuTHER

**Table 1. Lipid-Associated CpG Sites Replicated in Independent Cohorts With Whole Blood or CD4<sup>+</sup> T Cells DNA Methylation**

Lipid Trait	Type of Loci	CpG	Chr	Position	Gene Property	Gene	Direction*	Discovery $\beta$ (SE)	P Value	Direction†	Replication $\beta$ (SE)	P Value
Primary model												
TC	Novel	cg17901584	1	55353706	TSS1500	<i>DHCR24</i>	++	0.000149 (0.000027)	4.73E-08	+++	0.000114 (0.000032)	3.80E-04
TC	Novel	cg23759710	2	42990957	First exon	<i>OXER1</i>	--	-0.000082 (0.000014)	2.45E-09	---+	-0.000074 (0.000022)	8.32E-04
TC	Novel	cg00285394	8	126011954	Body	<i>SQLE</i>	++	0.000213 (0.000034)	5.98E-10	+++	0.000161 (0.000039)	3.16E-05
TC	Novel	cg07839457	16	57023022	TSS1500	<i>NLR5</i>	++	0.000273 (0.000051)	8.74E-08	+++	0.000231 (0.000049)	2.32E-06
TC	Novel	cg09978077	22	42229983	Body	<i>SREBF2</i>	++	0.000057 (0.000010)	1.79E-09	+++	0.000047 (0.000014)	4.82E-04
LDL-C	Novel	cg00285394	8	126011954	Body	<i>SQLE</i>	++	0.000230 (0.000039)	4.12E-09	++	0.000200 (0.000045)	1.04E-05
HDL-C	Novel	cg17901584	1	55353706	TSS1500	<i>DHCR24</i>	++	0.000717 (0.000052)	5.47E-43	++	0.000321 (0.000090)	3.55E-04
HDL-C	Novel	cg07567724	1	153777721	3'UTR	<i>GATAD2B</i>	--	-0.000248 (0.000045)	3.81E-08	--	-0.000311 (0.000087)	3.53E-04
HDL-C	Novel	cg19351166	2	209133632	5'UTR	<i>PIKFYVE</i>	--	-0.000207 (0.000038)	4.16E-08	--	-0.000279 (0.000065)	1.96E-05
HDL-C	Novel	cg06560379	6	44231305	Body	<i>NFKBIE</i>	++	0.000114 (0.000017)	6.05E-11	++	0.000083 (0.000023)	2.76E-04
HDL-C	Novel	cg16407699	10	74020428	...	...	--	-0.000280 (0.000048)	7.30E-09	--	-0.000265 (0.000069)	1.09E-04
HDL-C	Novel	cg19750657	13	38935967	3'UTR	<i>UFM1</i>	--	-0.000308 (0.000049)	2.24E-10	--	-0.000329 (0.000063)	2.20E-07
HDL-C	Novel	cg07814318	15	31624584	Body	<i>KLF13</i>	--	-0.000322 (0.000049)	6.38E-11	--	-0.000455 (0.000121)	1.69E-04
HDL-C	Novel	cg06192883	15	52554171	Body	<i>MYO5C</i>	--	-0.000258 (0.000043)	2.04E-09	--	-0.000283 (0.000053)	8.95E-08
HDL-C	Novel	cg11024682	17	17730094	Body	<i>SREBF1</i>	--	-0.000213 (0.000036)	4.19E-09	--	-0.000229 (0.000058)	8.72E-05
HDL-C	Previous	cg06500161	21	43656587	Body	<i>ABCG1</i>	--	-0.000459 (0.000037)	1.20E-34	--	-0.000322 (0.000072)	7.43E-06
HDL-C	Novel	cg06397161	22	39760059	5'UTR/Body	<i>SPRY4</i>	--	-0.000259 (0.000048)	7.25E-08	--	-0.000327 (0.000075)	1.27E-05
TG	Novel	cg03725309	1	109757585	Body	<i>SARS</i>	--	-0.009052 (0.001209)	7.11E-14	---	-0.011098 (0.001746)	2.09E-10
TG	Novel	cg16246545	1	120255941	Body	<i>PHGDH</i>	--	-0.012602 (0.002184)	7.94E-09	---	-0.007394 (0.002158)	6.11E-04
TG	Novel	cg14476101	1	120255992	Body	<i>PHGDH</i>	--	-0.021504 (0.002789)	1.25E-14	---	-0.012696 (0.002473)	2.82E-07
TG	Previous	cg19693031	1	145441552	3'UTR	<i>TXNIP</i>	--	-0.017424 (0.002249)	9.40E-15	---	-0.010932 (0.002263)	1.35E-06
TG	Novel	cg06690548	4	139162808	Body	<i>SLC7A11</i>	--	-0.021855 (0.002394)	6.98E-20	---	-0.007707 (0.001666)	3.70E-06
TG	Novel	cg21429551	7	30635762	Body	<i>GARS</i>	--	-0.026953 (0.003214)	4.97E-17	---	-0.012744 (0.002780)	4.57E-06
TG	Novel	cg03068497	7	30635838	Body	<i>GARS</i>	--	-0.025125 (0.003427)	2.27E-13	---	-0.012680 (0.002935)	1.55E-05

(Continued)

Table 1. Continued

Lipid Trait	Type of Loci	CpG	Chr	Position	Gene Property	Gene	Direction*	Discovery $\beta$ (SE)	P Value	Direction†	Replication $\beta$ (SE)	P Value
TG	Novel	cg19390658	7	30636176	Body	<i>GARS</i>	---	-0.020977 (0.002685)	5.61E-15	---	-0.012653 (0.002450)	2.40E-07
TG	Previous	cg07504977	10	102131012	...	...	++	0.012564 (0.002025)	5.45E-10	+++	0.011962 (0.001901)	3.10E-10
TG	Previous	cg00574958	11	68607622	5'UTR	<i>CPT1A</i>	---	-0.008999 (0.000724)	1.65E-35	+-	-0.011979 (0.001126)	2.01E-26
TG	Previous	cg09737197	11	68607675	5'UTR	<i>CPT1A</i>	---	-0.007154 (0.001104)	9.13E-11	+-	-0.010376 (0.001785)	6.18E-09
TG	Previous	cg17058475	11	68607737	5'UTR	<i>CPT1A</i>	---	-0.009858 (0.001003)	8.33E-23	---	-0.013425 (0.001665)	7.55E-16
TG	Novel	cg08129017	17	17728660	Body	<i>SREBF1</i>	++	0.009755 (0.001627)	2.02E-09	+++	0.011616 (0.002263)	2.84E-07
TG	Previous	cg11024682	17	17730094	Body	<i>SREBF1</i>	++	0.010107 (0.001327)	2.59E-14	+++	0.010284 (0.001345)	2.07E-14
TG	Novel	cg08857797	17	40927699	Body	<i>VPS25</i>	++	0.009577 (0.001708)	2.06E-08	+++	0.007994 (0.001849)	1.53E-05
TG	Novel	cg02711608	19	47287964	5'UTR/Body	<i>SLC1A5</i>	---	-0.008321 (0.001310)	2.14E-10	---	-0.004606 (0.001200)	1.24E-04
TG	Previous	cg27243685	21	43642366	5'UTR/Body	<i>ABCG1</i>	++	0.012223 (0.001164)	8.12E-26	+++	0.004508 (0.001093)	3.72E-05
TG	Novel	cg01176028	21	43653234	Body	<i>ABCG1</i>	++	0.006953 (0.001189)	5.00E-09	+++	0.011307 (0.002056)	3.82E-08
TG	Previous	cg06500161	21	43656587	Body	<i>ABCG1</i>	++	0.019854 (0.001359)	2.29E-48	+++	0.012731 (0.001637)	7.55E-15
Secondary BMI-adjusted model												
TC	Novel	cg23759710	2	42990957	First exon	<i>OXER1</i>	---	-0.000081 (0.000014)	3.60E-09	---	-0.000075 (0.000022)	7.75E-04
TC	Novel	cg00285394	8	126011954	Body	<i>SQLE</i>	++	0.000215 (0.000035)	4.83E-10	+++	0.000155 (0.000039)	6.08E-05
TC	Novel	cg07839457	16	57023022	TSS1500	<i>NLRC5</i>	++	0.000272 (0.000051)	9.62E-08	+++	0.000223 (0.000049)	5.54E-06
TC	Novel	cg09978077	22	42229983	Body	<i>SREBF2</i>	++	0.000057 (0.000010)	3.03E-09	+++	0.000047 (0.000014)	6.52E-04
LDL-C	Novel	cg00285394	8	126011954	Body	<i>SQLE</i>	++	0.000234 (0.000039)	2.32E-09	++	0.000197 (0.000046)	1.82E-05
HDL-C	Novel	cg19273683	1	21656047	Body	<i>ECE1</i>	---	-0.000289 (0.000051)	1.50E-08	---	-0.000244 (0.000078)	1.71E-03
HDL-C	Previous	cg06500161	21	43656587	Body	<i>ABCG1</i>	---	-0.000363 (0.000039)	2.48E-20	---	-0.000219 (0.000075)	3.41E-03
TG	Novel	cg03725309	1	109757585	Body	<i>SARS</i>	---	-0.007844 (0.001258)	4.56E-10	---	-0.009679 (0.001832)	1.27E-07
TG	Novel	cg14476101	1	120255992	Body	<i>PHGDH</i>	---	-0.018779 (0.002897)	8.96E-11	---	-0.009380 (0.002574)	2.68E-04
TG	Previous	cg19693031	1	145441552	3'UTR	<i>TXNIP</i>	---	-0.016514 (0.002340)	1.70E-12	---	-0.010270 (0.002376)	1.54E-05
TG	Novel	cg06690548	4	139162808	Body	<i>SLC7A11</i>	---	-0.019833 (0.002487)	1.54E-15	---	-0.006180 (0.001744)	3.95E-04
TG	Novel	cg21429551	7	30635762	Body	<i>GARS</i>	---	-0.023803 (0.003349)	1.18E-12	---	-0.009124 (0.002903)	1.67E-03

(Continued)

Table 1. Continued

Lipid Trait	Type of Loci	CpG	Chr	Position	Gene Property	Gene	Direction*	Discovery $\beta$ (SE)	PValue	Direction†	Replication $\beta$ (SE)	PValue
TG	Novel	cg19390658	7	30636176	Body	<i>GARS</i>	---	-0.020864 (0.002797)	8.73E-14	---	-0.010873 (0.002568)	2.30E-05
TG	Previous	cg00574958	11	68607622	5'UTR	<i>CPT1A</i>	---	-0.007903 (0.000750)	5.81E-26	---	-0.010401 (0.001177)	9.59E-19
TG	Previous	cg09737197	11	68607675	5'UTR	<i>CPT1A</i>	---	-0.007034 (0.001152)	1.01E-09	---	-0.010243 (0.001874)	4.58E-08
TG	Previous	cg17058475	11	68607737	5'UTR	<i>CPT1A</i>	---	-0.009144 (0.001046)	2.24E-18	---	-0.012478 (0.001746)	8.93E-13
TG	Novel	cg08129017	17	17728660	Body	<i>SREBF1</i>	++	0.009346 (0.001693)	3.36E-08	+++	0.010037 (0.002373)	2.85E-05
TG	Previous	cg27243685	21	43642366	5'UTR/Body	<i>ABCG1</i>	++	0.010416 (0.001204)	5.08E-18	+++	0.004038 (0.001146)	4.26E-04
TG	Previous	cg06500161	21	43656587	Body	<i>ABCG1</i>	++	0.016873 (0.001396)	1.29E-33	+++	0.010841 (0.001712)	1.10E-06

BMI indicates body mass index; Chr, chromosome; CpG, cytosine-guanine dinucleotide; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol; and TG, triglyceride.

\*Direction of effect in Framingham Heart Study and Prospective Investigation of the Vasculature in Uppsala Seniors Study.

†Direction of effect in LBC1936 (Lothian Birth Cohorts of 1936), LBC1921, and GOLDN (Genetics of Lipid Lowering Drugs and Diet Network) for TC and TG, and direction of effect in LBC1936 and GOLDN for LDL-C and HDL-C.

study<sup>25</sup> (Tables II through IX in the [Data Supplement](#)). Less than half of HDL-C (40%) and triglyceride-associated (46%) sites were associated in SAT, and more than half of HDL-C sites were in opposite directions in blood and adipose tissue, indicating that there may be independent regulatory effects across tissue types.

In addition, fixed effects meta-analyses across all 5 cohorts were performed for each lipid trait, identifying additional differentially methylated candidate regions that may play a role in lipid levels (Tables XI and XII in the [Data Supplement](#)), but that carry lesser weight given the lack of independent replication. Using the results of these meta-analyses, we investigated whether methylation at 15 CpGs associated with lipids in 2 recent publications<sup>38,39</sup> also was associated with the same lipid traits in our study. We found 12 (80%) CpGs reported in previous studies to be associated with the same lipid traits in our study (Table 2), highlighting the high degree of between-study replicability of lipid-methylation associations. Interestingly, the intergenic CpG cg07504977 associated with triglycerides in both our study and the previous study lies in an active regulatory region (DNase I hypersensitivity site and H3K27Ac mark) <10 kb distal to stearoyl-CoA desaturase (delta-9-desaturase). This gene plays an important role in the metabolism of dietary saturated fatty acids, a function that is critical for triglycerides synthesis and that has been shown to be disturbed in metabolic disease.<sup>40</sup> However, in our study, methylation at cg07504977 was not associated with expression of stearoyl-CoA desaturase in whole blood.

Many lipid-associated CpGs in our study were annotated to genes in loci highlighted in GWAS of cardiovascular traits, including lipids (*AMPD3*, *APOB*, *FADS2*, *GALNT2*, *LDLR*, *MYLIP*, and *TRIB1*), waist:hip ratio (*CBX3*, *KLF13*, and *LY86*), BMI (*ADCY3*), adiponectin (*TRIB1*), type 2 diabetes mellitus (*PTPRD*), and CHD (*APOB* and *LDLR*; Table XIII in the [Data Supplement](#)).

### Functional Annotation of Lipid-Associated CpGs

We explored the functional roles of the 193 CpGs associated with lipid traits by investigating their genomic locations with respect to genes, CpG islands, and functional regulatory elements. Lipid-associated CpGs were less commonly located in CpG islands ( $P=1.01E-15$ ) and promoters ( $P=5.82E-04$ ), when compared with all CpGs on the array (Figure XV in the [Data Supplement](#)). The observation that differential DNA methylation in relation to chronic human disease traits (as opposed to cancer) is less likely to be seen at promoters and CpG islands has been previously reported.<sup>25</sup> To further explore the regulatory activity of identified loci, we examined the overlap of the 193 lipid-associated CpGs with functional regulatory elements across cell types using RegulomeDB.<sup>41</sup> About 14% of sites showed strong evidence of being located in a functional regulatory region (RegulomeDB score 1a-2c; Tables II through IX in the [Data Supplement](#)); this was not more than expected by chance ( $P=0.83$ ).

To further the *in silico* identification of relevant affected tissues, we used the eFORGE tool (<http://eforge.cs.ucl.ac.uk/>), which determines whether the identified CpGs are enriched in DNase I hypersensitivity site hotspots in specific tissue types (Figure XVI in the [Data Supplement](#)). Our identified CpGs were in active DNase I hypersensitivity site hotspots across a range of tissue types in ENCODE and Epigenome Roadmap Consortium tissue sets (FDR  $Q$  value <0.01), specifically blood, liver, muscle, heart, and epithelium (adipose tissue is not represented in this tool). Notably, the identified CpGs were not in DNase I hypersensitivity site hotspots in nervous tissue (brain, cerebellum, hippocampus, and nervous), gastrointestinal tissue (colon, kidney, pancreas, and pancreatic duct), bone tissue, and eye tissue.

To place our findings in a broader biological context, we performed gene set enrichment analysis<sup>31,32</sup> for genes annotated to the 193 CpGs associated with lipid levels. For TC,



**Table 2. Associations of Lipid Levels With Methylation at CpGs Previously Reported to be Associated with Lipids**

Trait	References	CpG	Gene	$\beta$ (replicating)	Meta-Analysis Across All 5 Cohorts in Our Study	
					Direction	P Value*
LDL-C	Pfeiffer et al <sup>39</sup>	cg22178392	<i>TNIP1</i>	0.040 (yes)	+	1.33E-04
HDL-C	Pfeiffer et al <sup>39</sup>	cg06500161	<i>ABCG1</i>	−0.049 (yes)	−*	2.09E-38*
TG	Pfeiffer et al <sup>39</sup>	cg06500161	<i>ABCG1</i>	0.070 (yes)	+	4.20E-59*
TG	Irvin et al <sup>38</sup> ; Pfeiffer et al <sup>39</sup>	cg00574958	<i>CPT1A</i>	−0.118 (yes), −0.032	−*	7.22E-09*
TG	Irvin et al <sup>38</sup>	cg17058475	<i>CPT1A</i>	−0.035	−*	2.72E-36*
TG	Irvin et al <sup>38</sup>	cg09737197	<i>CPT1A</i>	−0.027	−*	1.05E-17*
TG	Irvin et al <sup>38</sup>	cg01082498	<i>CPT1A</i>	−0.011	−*	2.07E-13*
TG	Pfeiffer et al <sup>39</sup>	cg27243685	<i>ABCG1</i>	0.064 (yes)	+	2.00E-24*
TG	Pfeiffer et al <sup>39</sup>	cg19693031	<i>TXNIP</i>	−0.030 (yes)	−*	5.54E-19*
TG	Pfeiffer et al <sup>39</sup>	cg11024682	<i>SREBF1</i>	0.059 (yes)	+	3.74E-27*
TG	Pfeiffer et al <sup>39</sup>	cg07504977	...	0.026 (yes)	+	9.91E-19*
TG	Pfeiffer et al <sup>39</sup>	cg20544516	<i>MIR33B/SREBF1</i>	0.043 (yes)	+	3.57E-08*
TG	Pfeiffer et al <sup>39</sup>	cg07397296	<i>ABCG1</i>	0.027 (yes)	+	1.08E-10*
TG	Pfeiffer et al <sup>39</sup>	cg07815238	...	0.048 (no)	+	0.037
TG	Pfeiffer et al <sup>39</sup>	cg12556569†	<i>APOA5</i>	0.005 (no)	N/A	N/A

CpG indicates cytosine–guanine dinucleotide; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol; and TG, triglyceride.

\*CpGs associate with the same lipid trait in combined meta-analysis ( $P < 1.08E-7$ ) in our study.

†CpG excluded in our analysis because of common genetic variant 1 base away from CpG site.

the pathway analyses revealed enrichment in processes relating to sterol, lipid, and cholesterol metabolism and biosynthesis (FDR=0.0029–0.037), indicating that DNA methylation sites associated with cholesterol primarily affect processes directly relating to lipid production and metabolism (Tables XIV through XVI in the [Data Supplement](#)). For triglycerides, the pattern was different because metabolism of amino acids was highlighted in the pathway analyses (FDR=0.034). No significant enrichment in pathways was observed in analysis of genes annotated to CpG sites associated with LDL-C or HDL-C. When restricting the enrichment analyses to genes annotated to replicating CpGs or to those where methylation levels were associated with gene expression of their respective genes, we observed similar results (Tables XIV through XVI in the [Data Supplement](#)), with the exception of HDL-C, which now showed significant enrichment in lipid metabolism (FDR=0.0056–0.04).

### Genetic Regulation of Lipid-Associated DNA Methylation

To assess the role of genetic variation in controlling lipid-related DNA methylation changes, we studied the association of sequence variants in *cis* with methylation levels at lipid-associated CpGs (*cis*-meQTLs). Mapping of *cis*-meQTLs (SNPs in a 100 kb window around CpG sites) was performed in the FHS cohort (n=2246) with subsequent replication of lead meQTLs in the PIVUS cohort (n=775). In agreement with previous studies,<sup>6,25</sup> we found a large proportion of CpG sites to associate with common SNPs in *cis*. We found 123 out of 193 (64%) lipid-associated CpG sites to be at least partly

regulated by genetic sequence variation in *cis* ( $P < 1E-04$ ); 60 of these replicated in PIVUS (at  $P < 4.071E-04$ ; Table XVII in the [Data Supplement](#)).

We investigated whether the 123 significant lead meQTL SNPs or their proxies ( $r^2 > 0.8$ ) were over-represented among SNPs with nominally significant associations ( $P < 0.05$ ) in GWAS meta-analyses from the CARDIoGRAM consortium for CHD<sup>42</sup> and the Global Lipids Genetics consortium for lipid levels.<sup>4</sup> We found evidence of enrichment (applying a 1-sided Fisher exact test) of nominally significant associations for CHD ( $P = 7.04E-4$ ), TC ( $P = 4.36E-3$ ), HDL-C ( $P = 8.3E-3$ ), and triglycerides ( $P = 2.9E-5$ ) among the *cis*-meQTL lead SNPs (or proxies). Furthermore, we found the lead *cis*-meQTL SNP (rs563290) of cg05337441 (associated with LDL-C in discovery,  $P = 4.5E-8$  but not surviving Bonferroni cutoff threshold in replication,  $P = 1.7E-2$ ), located in an intron of *APOB*, to be associated with LDL-C in GWAS<sup>43</sup> and to be a highly correlated proxy ( $r^2 = 1$ ) of genome-wide significant GWAS index SNPs (rs151535 and rs562338; located  $\approx 20$  kb upstream of the *APOB* transcription start site) in LDL-C meta-analyses.<sup>44,45</sup> This *cis*-meQTL proxy for *APOB* locus methylation (rs151535) is also associated with CHD at a genome-wide level of significance ( $P = 1.8E-10$ ) from the CardiogramC4D consortium data.<sup>42</sup>

### The Impact of Lipid-Associated CpGs on Gene Expression

Examining gene expression in relation to DNA methylation in blood from participants in the FHS, we investigated whether methylation levels at lipid-associated CpGs were

associated with mRNA expression levels of nearby genes ( $\pm 500$  kb). We found 29 CpGs (out of 193 tested; 15%) to be associated with expression in blood of at least 1 adjacent gene ( $FDR < 0.05$ ; 36 CpG–expression pairs in total; Table XVIII in the [Data Supplement](#)). For the majority (86%) of these associations, levels of methylation and expression were inversely correlated. For 17 of these 29 CpGs (59%), there was also a significant *cis*-meQTL. The lead meQTL SNP was significantly associated with both methylation and gene expression ( $FDR < 0.05$ ) for 12 of 36 CpG–expression pairs (29 unique CpGs), suggesting that the genotype may affect both methylation and expression. This was the case for the following genes: *CHSY1* (cg24002003), *DHCR24* (cg17901584), *ECE1* (cg19273683), *IL18R1* (cg05295703), *IL1RL1* (cg05295703), *KANK2* (cg01751802), *LDLR* (cg26313301), *PHGDH* (cg14476101, cg16246545), *PRKD2* (cg22304262), *SREBF1* (cg08129017), and *SREBF2* (cg09978077). For the remaining 6 CpG–expression pairs, the meQTL SNP was associated with methylation ( $FDR < 0.05$ ) but not with expression ( $FDR > 0.05$ ) as presented in Table XVIII in the [Data Supplement](#).

### Detailed Characterization of Lipid CpG Sites Using Metabolomics

To further characterize functional relevance of lipid-associated CpG sites, we tested levels of methylation at the 193 CpGs for association with 229 serum metabolites in the PIVUS cohort.<sup>46</sup> We found 29 of the lipid-associated CpGs to be associated with at least 1 metabolite ( $FDR < 0.01$ ; Table XIX in the [Data Supplement](#)). As expected, the majority of the associations were between a lipid-related CpG site and various lipid-derived metabolites (Figure XVII in the [Data Supplement](#)). Most associations were observed with cg17901584 in the promoter of *DHCR24* (associated with TC, HDL-C, and triglycerides) and with sites in the promoter of *ABCG1* (associated with HDL-C and triglycerides), highlighting the central role for these genes in lipid metabolism. Metabolites associated with methylation of the *DHCR24* promoter included a derivative of cinnamic acid, recently shown to be associated with a lower risk of incident CHD events.<sup>46</sup> Methylation at the *ABCG1* locus was associated with specific ceramides and sphingomyelins, which have been implicated in the development of atherosclerosis and CHD.<sup>47,48</sup>

### Association of Lipid-Associated CpGs With Disease Outcomes

We investigated whether the 33 replicating lipid-associated CpG sites were also associated with incident CHD events during an 8-year follow-up in the FHS (number of CHD events=115) and a 10-year follow-up in PIVUS (number of CHD events=78) using multivariable Cox proportional hazard models. Methylation levels at *ABCG1* (cg27243685) were significantly associated (Bonferroni-corrected  $\alpha < 0.05$ , nominal  $P < 1.52E-03$ ) with CHD in a meta-analysis of FHS and PIVUS (hazard ratio per SD increment=1.38; 95% confidence interval, 1.15–1.66;  $P=6.86E-04$ ; Table XX in the [Data Supplement](#)). We found the relationship of methylation at cg27243685 with triglycerides and risk of CHD to be directionally consistent with the expected based on previous

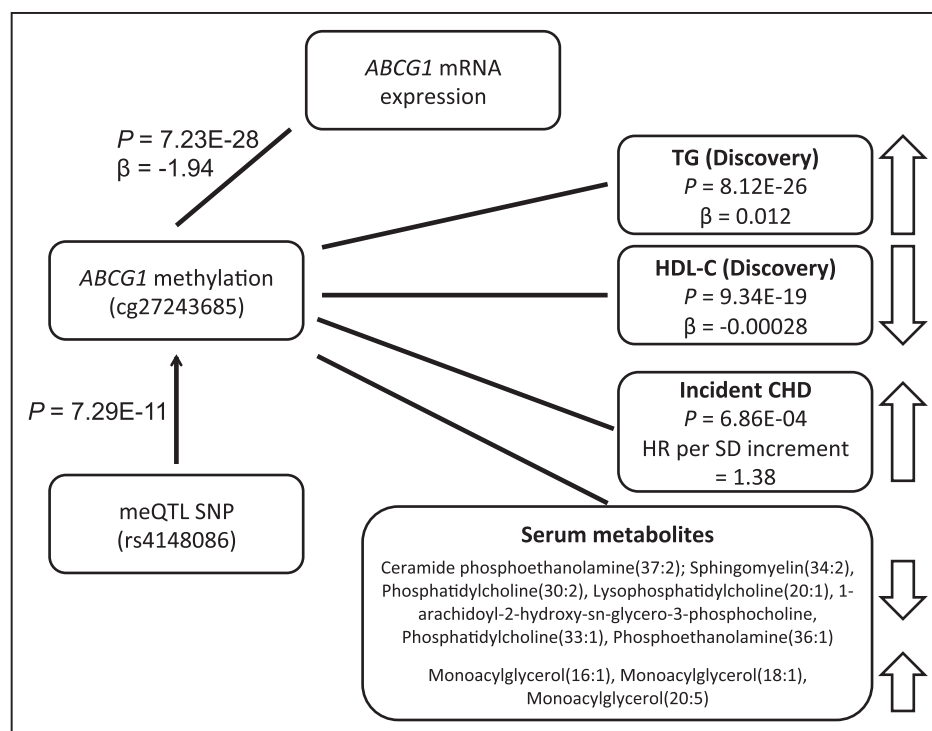
studies of lipid levels and CHD risk.<sup>1–3</sup> Hypermethylation at cg27243685 in the 5′-UTR of *ABCG1*—that was associated with decreased expression of *ABCG1* (Table XVIII in the [Data Supplement](#))—was associated with higher triglycerides and lower HDL-C, as well as increased risk for CHD (Figure 2). This *ABCG1* locus (cg27243685) was also highlighted in the previous sections as containing a *cis*-meQTL and being associated with metabolites. This illustrates an example of a pathway linking genetic variant to perturbed DNA methylation, altered expression levels, circulating metabolites, lipid levels (triglycerides and HDL-C), and risk of CHD (Figure 2).

## Discussion

In this study, we aimed to identify epigenetic variation associated with serum lipid concentrations, which are among the most established risk factors for CVD. We report findings of a genome-wide scan of blood DNA methylation in relation to circulating lipid levels from  $\leq 2306$  individuals with independent external replication in  $\leq 2025$  additional individuals. We extend the findings of published literature on the association of differential DNA methylation with circulating lipids<sup>38,39,49,50</sup> by examining larger discovery and replication samples and by examining the association of methylation at the associated CpGs with gene expression, intermediate metabolites, and incident CHD. We have made several novel observations about the role of DNA methylation in the regulation of lipids and risk of CVD and highlight 3 important contributions. First, we identified novel replicated loci of differential methylation in blood associated with circulating lipid levels that may represent potential therapeutic targets. Second, we describe the overlap of methylation and GWAS SNPs and identify a potential mechanism of a known LDL-C–related GWAS variant at the *APOB* locus acting as a *cis*-meQTL on LDL-C–related differential methylation at cg05337441, intronic to *APOB*. Third, we identify HDL-C–related and triglyceride-related differential methylation at the *ABCG1* locus (cg27243685) to be associated with expression of a gene involved in reverse cholesterol transport (*ABCG1*), metabolites that influence reverse cholesterol transport (sphingomyelins), and subsequently to be associated with a 38% higher risk of incident CHD per SD increase in methylation.

We found methylation at 193 CpG sites to be associated with lipid levels and replicated 33 of these in 3 independent cohorts with data on DNA methylation in blood and T cells. Many of the differentially methylated loci associated with LDL-C, triglycerides, and to a lesser degree HDL-C, were independent of adjustment for BMI. Twenty-five of the 33 replicated CpGs have not been previously reported to be associated with lipid levels.<sup>38,39</sup> Novel sites included those near genes with a known function in cholesterol metabolism (*DHCR24*, *SREBF2*, and *SQLE*) and with a possible role in atherosclerosis (endothelin-converting enzyme-1).<sup>51,52</sup> The novel genes identified warrant further research as potential targets for perturbation to reduce dyslipidemia.

When exploring whether methylation at lipid-associated CpGs has also been associated with related cardiometabolic traits, we found overlap with associations for adiposity (near genes *ABCG1*, *CPT1A*, *DHCR24*, *KLF13*, *MYO5C*, *PHGDH*, *SREBF1*, and *VPS25*),<sup>53–55</sup> glycemic traits (near *ABCG1*),<sup>56</sup>



**Figure 2.** Associations at the *ABCG1* locus. CHD indicates coronary heart disease; HDL-C, high-density lipoprotein cholesterol; meQTL, methylation quantitative trait locus; SNP, single-nucleotide polymorphism; and TG, triglyceride.

and type 2 diabetes mellitus (near genes *SREBF1*, *ABCG1*, and *TXNIP*).<sup>57</sup> In addition, we observed associations of circulating lipids with DNA methylation levels at CpGs near genes previously reported to be associated with lipids, other cardiovascular traits, and CVD events in GWAS.

Further, pathway analyses, including genes annotated to lipid-associated CpGs, showed enrichment in pathways involved in lipid, sterol, and cholesterol metabolic and biosynthesis processes for cholesterol-related CpGs, whereas amino acid metabolism pathways were enriched for triglyceride-associated CpGs. These observations highlight the different biological mechanisms underlying changes in genomic regulation observed in association with TC and TGs.

We identified genetic drivers of lipid-associated CpGs in blood through integration with SNPs in *cis*-meQTLs analyses. At 64% of the lipid-associated CpGs, the effect is determined in part by genotype. GWAS SNPs for lipids and CHD were enriched among the *cis*-meQTL SNPs of lipid-associated CpGs. Further, we observed association with expression levels of adjacent genes for 15% of the CpGs, which indicates possible mechanisms of effect through changes in transcription. For 17 of the lipid-related CpGs where there was an association with expression levels of an adjacent gene, there was also a significant *cis*-meQTL. For the majority of these, the genotype affected both methylation and gene expression. In these instances, our data provide evidence linking multiple steps from genetic variants affecting DNA methylation, to modulation of gene expression to effects on circulating lipid levels. For example, at the *ABCG1* locus, we observed that the minor allele at intronic variant rs4148086 was associated with increased methylation at cg27243685. This methylation marker, which is located at the south shelf of a CpG island in the 5'-UTR region

of *ABCG1*, was associated with decreased expression of *ABCG1* in blood, increased triglyceride levels (even after adjustment of BMI and regulated both by blood and SAT methylation), and increased risk of new-onset CHD. Methylation in this locus (at cg06500161) has previously been associated with prevalent myocardial infarction.<sup>39</sup> The *ABCG1* gene product functions in the efflux of cholesterol from lipid-loaded macrophages to HDL-C.<sup>58</sup> However, the functional basis for association to levels of triglycerides in blood circulation is unclear. Although circulating HDL-C levels has been largely disproven as a causal factor for CHD,<sup>2</sup> the importance of cholesterol efflux function in CHD risk is an emerging topic of discussion.<sup>59</sup> In addition to cholesterol, *ABCG1* mediates the efflux of sphingomyelin and phosphatidylcholine, and the cholesterol efflux by *ABCG1* has been demonstrated to have some dependence on sphingomyelin concentrations.<sup>60,61</sup> Sphingomyelins have been implicated in the development of atherosclerosis and CHD.<sup>48,62</sup> In our study, methylation in the *ABCG1* locus was also associated with specific sphingomyelins and ceramides (also implicated in CHD<sup>47</sup>). Methylation at CpG sites in the *ABCG1*, as well as the *DHCR24* loci, was also associated with a large number of other lipid-related metabolites in blood, further highlighting the central role for these genes in processes relating to lipid metabolism and development of CVD.

The main strengths of this study include the large sample size of the genome-wide DNA methylation and  $\leq 10$  years of follow-up allowing analyses of incident CHD end points. In addition, inclusion of several other types of functional genomics data (gene expression and metabolites) helped us to draw more precise conclusions on the links between methylation and circulating lipid levels. We replicated a large fraction of previously reported associations of methylation and lipid levels,



providing assurance that associations of methylation with lipid levels are reliable across different studies and indicate that also the novel findings reported may indeed represent true findings.

The study also has limitations. Blood-derived cells, although easily accessible and good for biomarker discovery, may not be the most relevant tissue for drawing biological conclusions. Our validation in adipose tissue reveals that at least a proportion of the observed associations are shared across tissues. The cross-sectional design does not allow us to determine the causal relationship between lipid and DNA methylation. Our analysis of lipid-associated CpGs with incident disease indicates the relevance of methylation in at least one of these CpGs for disease pathophysiology. Further, a relatively low proportion of our findings could be robustly validated in the replication stage. However, it should be noted that we observed a high level of agreement of  $\beta$  coefficients even for CpGs that did not formally replicate at the  $P$  value threshold. This indicates that the low replication rate may be because of smaller sample size in the replication stage, particularly for LDL-C and HDL-C, giving reduced power, especially in the light of our strict criteria for replication (which was chosen to minimize false-positive findings). In addition, if the differentially methylated CpGs identified in discovery from whole blood did not also occur in CD4<sup>+</sup> T cells, we would not expect to see replication in the GOLDN replication cohort that assayed DNA from cell-sorted CD4<sup>+</sup> T cells. Furthermore, cholesterol panels from the LBC cohort were obtained in a nonfasting state and may have reduced our ability to replicate findings. Finally, transcriptomic and metabolomic data were not available in every cohort, and, therefore, we were not able to demonstrate similar findings in each participating study.

In conclusion, we report novel associations of DNA methylation with lipid levels. We identify links between genetic variation underlying lipids and CHD to differential DNA methylation. We also highlight HDL-C-related and triglyceride-related differential methylation and expression of a reverse cholesterol transporter, *ABCG1*, and the association with an increased risk of incident CHD. Our findings highlight established and novel targets and mechanisms that can be used as a starting point for potential new treatments for dyslipidemia and CVD.

### Acknowledgments

The FHS (Framingham Heart Study), PIVUS (Prospective Investigation of the Vasculature in Uppsala Seniors Study), LBC (Lothian Birth Cohorts), and GOLDN (Genetics of Lipid Lowering Drugs and Diet Network) studies thank the cohort participants and team members who contributed to these studies. Data on coronary artery disease/myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from <http://www.cardiogramplusc4d.org/>. The views expressed in this article are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute (NHLBI); the National Institutes of Health (NIH) or the US Department of Health and Human Services.

### Sources of Funding

FHS (Framingham Heart Study) is funded by National Institutes of Health (NIH) contract N01-HC-25195 and HHSN268201500001I and administered by Boston University. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute (NHLBI), NIH, and an

NIH Director's Challenge Award (Dr Levy, Principal Investigator). The analytic component of this project was funded by the Division of Intramural Research, NHLBI, and the Center for Information Technology, NIH, Bethesda, MD. This study used the computational resources of the Biowulf system at the NIH, Bethesda, MD (<https://hpc.nih.gov/>). Dr Mendelson is partly supported by the Tommy Kaplan Fund, Boston Children's Hospital. Dr Liang is partially supported by NIH grant P30 DK46200. Dr Ingelsson is supported by Knut and Alice Wallenberg (KAW) Foundation, Swedish Research Council (VR; grant no. 2012-1397), Swedish Heart-Lung Foundation (20120197) NIH grants 1R01DK106236-01A1 and 1R01HL135313-01. Genome-wide DNA methylation profiling in PIVUS was funded by the Uppsala University Hospital (ALF-medel) and was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the VR and the KAW Foundation. Phenotype collection in the LBC1921 (Lothian Birth Cohorts of 1921) study was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the LBC1936 (Lothian Birth Cohorts of 1936) study was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE; Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. Drs Marioni, Starr, and Deary are members of the University of Edinburgh CCACE. CCACE is supported by funding from the BBSRC, the Medical Research Council and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). Research reported in this publication was supported by National Health and Medical Research Council (NHMRC) project grant 1010374 and an NHMRC Fellowship to Dr McRae (1083656). The GOLDN (Genetics of Lipid Lowering Drugs and Diet Network) study was supported by NIH National Heart, Lung and Blood Institute grant R01 HL104135-01. The MuTHER study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013). The study as part of TwinsUK also receives support from the Medical Research Council, European Union, National Institute for Health Research-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. Dr Spector is a holder of an European Research Council Advanced Principal Investigator award.

### Disclosures

Erik Ingelsson is an advisor and consultant for Precision Wellness, Inc., and advisor for Cellink for work unrelated to the present project.

### Appendix

From the Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory (Å.K.H., S.G., E.I.) and Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory (J.K.S.), Uppsala University, Sweden; Cardiovascular Medicine unit, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden (Å.K.H.) Framingham Heart Study, MA (M.M.M., R.J., C.Y., C.L., T.H., S.D., L.A.C., D.L.); Department of Biostatistics (C.L., L.A.C., S.D.), Boston University, MA; Boston University, MA (M.M.M.); Department of Cardiology, Boston Children's Hospital, MA (M.M.M.); Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD (M.M.M., R.J., C.Y., C.L., T.H., D.L.); Centre for Cognitive Ageing and Cognitive Epidemiology (R.E.M., J.M.S., I.J.D.), Medical Genetics Section, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine (R.E.M.), Alzheimer Scotland Dementia Research Centre (J.M.S.), and Department of Psychology (I.J.D.), University of Edinburgh, United Kingdom; Queensland Brain Institute, The University of Queensland, Brisbane, Australia (R.E.M., A.F.M., S.S.); Institute for Molecular Bioscience, University of



Queensland, Brisbane, Queensland, Australia (S.S., A.F.M.); Hebrew Senior Life, Harvard Medical School, Boston, MA (R.J.); Department of Epidemiology, School of Public Health (M.R.I.) and Department of Biostatistics, Section on Statistical Genetics (D.Z.), University of Alabama at Birmingham; Department of Biostatistics, Harvard School of Public Health, Boston, MA (L. Liang); William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, United Kingdom (P.D.); Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia (P.D.); Department of Twin Research and Genetic Epidemiology, King's College London, United Kingdom (T.D.S.); Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University Hospital, Sweden (J.S., L.L.); Children's Hospital Oakland Research Institute, CA (R.M.K.); College of Public Health, University of Kentucky, Lexington (D.K.A.); and Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, CA (E.I.).

## References

- Mendis S, Puska P, Norrving B. *Global Atlas on Cardiovascular Disease Prevention and Control*. Geneva, Switzerland: World Health Organization; 2011.
- Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet*. 2012;380:572–580. doi: 10.1016/S0140-6736(12)60312-2.
- Do R, Willer CJ, Schmidt EM, Sengupta S, Gao C, Peloso GM, et al. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*. 2013;45:1345–1352. doi: 10.1038/ng.2795.
- Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283.
- Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25–33.
- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. 2011;12:R10. doi: 10.1186/gb-2011-12-1-r10.
- Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci USA*. 2008;105:17046–17049. doi: 10.1073/pnas.0806560105.
- Teh AL, Pan H, Chen L, Ong ML, Dogra S, Wong J, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res*. 2014;24:1064–1074. doi: 10.1101/gr.171439.113.
- Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al; MuTHER Consortium. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012;8:e1002629. doi: 10.1371/journal.pgen.1002629.
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet*. 2013;22:843–851. doi: 10.1093/hmg/ddt488.
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12:529–541. doi: 10.1038/nrg3000.
- Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O, et al. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun*. 2016;7:11089. doi: 10.1038/ncomms11089.
- Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheimer J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci USA*. 2016;113:E1826–E1834. doi: 10.1073/pnas.1519286113.
- Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, et al. Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. *Schizophr Bull*. 2016;42:406–414. doi: 10.1093/schbul/sbv074.
- Hardy T, Zeybel M, Day CP, Dipper C, Masson S, McPherson S, et al. Plasma DNA methylation: a potential biomarker for stratification of liver fibrosis in non-alcoholic fatty liver disease [published online ahead of print March 21, 2016]. *Gut*. doi: 10.1136/gutjnl-2016-311526. Available at: <http://gut.bmj.com/content/early/2016/03/21/gutjnl-2016-311526.full>. Accessed April 15, 2016.
- Huang YT, Chu S, Loucks EB, Lin CL, Eaton CB, Buka SL, et al. Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. *Epigenetics*. 2016;11:227–236.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110:281–290.
- Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler Thromb Vasc Biol*. 2005;25:2368–2375. doi: 10.1161/01.ATV.0000184769.22061.da.
- Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J Pers Soc Psychol*. 2004;86:130–147. doi: 10.1037/0022-3514.86.1.130.
- Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr*. 2007;7:28. doi: 10.1186/1471-2318-7-28.
- Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2012;41:1576–1584. doi: 10.1093/ije/dyr197.
- Irvin MR, Kabagambe EK, Tiwari HK, Parnell LD, Straka RJ, Tsai M, et al. Apolipoprotein E polymorphisms and postprandial triglyceridemia before and after fenofibrate treatment in the Genetics of Lipid Lowering and Diet Network (GOLDN) Study. *Circ Cardiovasc Genet*. 2010;3:462–467. doi: 10.1161/CIRCGENETICS.110.950667.
- Zhang X, Johnson AD, Hendricks AE, Hwang SJ, Tanriverdi K, Ganesh SK, et al. Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes. *Hum Mol Genet*. 2014;23:782–795. doi: 10.1093/hmg/ddt461.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39:906–913. doi: 10.1038/ng2088.
- Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al; Multiple Tissue Human Expression Resource Consortium. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet*. 2013;93:876–890. doi: 10.1016/j.ajhg.2013.10.004.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98:288–295. doi: 10.1016/j.ygeno.2011.07.007.
- Vazquez AI, Bates DM, Rosa GJ, Gianola D, Weigel KA. Technical note: an R package for fitting generalized linear mixed models in animal breeding. *J Anim Sci*. 2010;88:497–504. doi: 10.2527/jas.2009-1952.
- Atkinson B, Therneau T. Kinship: mixed-effects cox models, sparse matrices, and modeling data from large pedigrees. R package version 1.1.0–17. 2007. Available at: <https://cran.r-project.org/web/packages/kinship/index.html>.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340.
- Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al; Multiple Tissue Human Expression Resource (MuTHER) Consortium. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012;44:1084–1089. doi: 10.1038/ng.2394.
- Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1–13.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44–57. doi: 10.1038/nprot.2008.211.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86. doi: 10.1186/1471-2105-13-86.
- Team RC. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015.

35. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363–1369. doi: 10.1093/bioinformatics/btu049.
36. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003;100:9440–9445. doi: 10.1073/pnas.1530509100.
37. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. 2007;3:1724–1735. doi: 10.1371/journal.pgen.0030161.
38. Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, et al. Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study. *Circulation*. 2014;130:565–572. doi: 10.1161/CIRCULATIONAHA.114.009158.
39. Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, et al. DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet*. 2015;8:334–342. doi: 10.1161/CIRCGENETICS.114.000804.
40. Merino DM, Ma DW, Mutch DM. Genetic variation in lipid desaturases and its impact on the development of human disease. *Lipids Health Dis*. 2010;9:63. doi: 10.1186/1476-511X-9-63.
41. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790–1797. doi: 10.1101/gr.137323.112.
42. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al; Cardiogenics; CARDIOGRAM Consortium. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*. 2011;43:333–338. doi: 10.1038/ng.784.
43. Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, et al; ASCOT Investigators; NORDIL Investigators; BRIGHT Consortium. Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. *Am J Hum Genet*. 2009;85:628–642. doi: 10.1016/j.ajhg.2009.10.014.
44. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*. 2009;41:56–65. doi: 10.1038/ng.291.
45. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*. 2008;40:161–169. doi: 10.1038/ng.76.
46. Ganna A, Salihovic S, Sundström J, Broeckling CD, Hedman AK, Magnusson PK, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet*. 2014;10:e1004801. doi: 10.1371/journal.pgen.1004801.
47. Bismuth J, Lin P, Yao Q, Chen C. Ceramide: a common pathway for atherosclerosis? *Atherosclerosis*. 2008;196:497–504. doi: 10.1016/j.atherosclerosis.2007.09.018.
48. Borodicz S, Czarzasta K, Kuch M, Cudnoch-Jedrzejewska A. Sphingolipids in cardiovascular diseases and metabolic disorders. *Lipids Health Dis*. 2015;14:55. doi: 10.1186/s12944-015-0053-y.
49. Frazier-Wood AC, Aslibekyan S, Absher DM, Hopkins PN, Sha J, Tsai MY, et al. Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. *J Lipid Res*. 2014;55:1324–1330. doi: 10.1194/jlr.M048504.
50. Guay SP, Légaré C, Brisson D, Mathieu P, Bossé Y, Gaudet D, et al. Epigenetic and genetic variations at the TNNT1 gene locus are associated with HDL-C levels and coronary artery disease. *Epigenomics*. 2016;8:359–371. doi: 10.2217/epi.15.120.
51. Ihling C, Szobath T, Bohrmann B, Brockhaus M, Schaefer HE, Loeffler BM. Coexpression of endothelin-converting enzyme-1 and endothelin-1 in different stages of human atherosclerosis. *Circulation*. 2001;104:864–869.
52. Martínez-Miguel P, Raach V, Zaragoza C, Valdivielso JM, Rodríguez-Puyol M, Rodríguez-Puyol D, et al. Endothelin-converting enzyme-1 increases in atherosclerotic mice: potential role of oxidized low density lipoproteins. *J Lipid Res*. 2009;50:364–375. doi: 10.1194/jlr.M800215-JLR200.
53. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity*. 2015;23:1493–1501. doi: 10.1002/oby.21111.
54. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet*. 2015;24:4464–4479. doi: 10.1093/hmg/ddv161.
55. Ollikainen M, Ismail K, Gervin K, Kyllönen A, Hakkarainen A, Lundbom J, et al. Genome-wide blood DNA methylation alterations at regulatory elements and heterochromatic regions in monozygotic twins discordant for obesity and liver fat. *Clin Epigenetics*. 2015;7:39. doi: 10.1186/s13148-015-0073-5.
56. Hidalgo B, Irvin MR, Sha J, Zhi D, Aslibekyan S, Absher D, et al. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. *Diabetes*. 2014;63:801–807. doi: 10.2337/db13-1100.
57. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol*. 2015;3:526–534. doi: 10.1016/S2213-8587(15)00127-8.
58. Smith JD. Insight into ABCG1-mediated cholesterol efflux. *Arterioscler Thromb Vasc Biol*. 2006;26:1198–1200. doi: 10.1161/01.ATV.0000221217.86465.66.
59. Khera AV, Cuchel M, de la Llera-Moya M, Rodríguez A, Burke MF, Jafri K, et al. Cholesterol efflux capacity, high-density lipoprotein function, and atherosclerosis. *N Engl J Med*. 2011;364:127–135. doi: 10.1056/NEJMoa1001689.
60. Sano O, Kobayashi A, Nagao K, Kumagai K, Kioka N, Hanada K, et al. Sphingomyelin-dependence of cholesterol efflux mediated by ABCG1. *J Lipid Res*. 2007;48:2377–2384. doi: 10.1194/jlr.M700139-JLR200.
61. Gulshan K, Brubaker G, Wang S, Hazen SL, Smith JD. Sphingomyelin depletion impairs anionic phospholipid inward translocation and induces cholesterol efflux. *J Biol Chem*. 2013;288:37166–37179. doi: 10.1074/jbc.M113.512244.
62. Liu J, Huan C, Chakraborty M, Zhang H, Lu D, Kuo MS, et al. Macrophage sphingomyelin synthase 2 deficiency decreases atherosclerosis in mice. *Circ Res*. 2009;105:295–303. doi: 10.1161/CIRCRESAHA.109.194613.

## CLINICAL PERSPECTIVE

Serum lipid levels are among the most established risk factors for cardiovascular disease, the leading cause of death globally. In this study, we report on the relations of circulating serum lipids with epigenetic marks and also provide evidence of a role for epigenetics in cardiovascular disease development. We present findings from a genome-wide scan of blood DNA methylation in relation to circulating lipid levels from 2306 individuals with independent external replication in 2025 individuals. We have made several novel observations about the role of DNA methylation in the regulation of lipids and risk of cardiovascular disease and highlight 3 important contributions: (1) we identify novel replicated loci of differential methylation in blood associated with circulating lipid levels that may represent potential therapeutic targets, (2) we describe the overlap of methylation and genome-wide association studies single-nucleotide polymorphisms and identify a potential mechanism of a known low-density lipoprotein cholesterol-related and coronary heart disease-related single-nucleotide polymorphism from genome-wide association studies at the *APOB* locus acting as a *cis*-methylation quantitative trait locus on low-density lipoprotein cholesterol-related differential methylation at a site intronic to *APOB*, and (3) we identify triglyceride- and high-density lipoprotein cholesterol-related differential methylation at the *ABCG1* locus to be associated with expression of a gene involved in reverse cholesterol transport (*ABCG1*), metabolites that influence reverse cholesterol transport (sphingomyelins), and subsequently to be associated with a 38% higher risk in incident coronary heart disease events. We think that our findings provide important insights into the contributions of epigenetics in circulating lipids and cardiovascular disease and may provide insights to novel therapeutic targets.

## Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease Events and Are Enriched for Results From Genome-Wide Association Studies

Åsa K. Hedman, Michael M. Mendelson, Riccardo E. Marioni, Stefan Gustafsson, Roby Joehanes, Marguerite R. Irvin, Degui Zhi, Johanna K. Sandling, Chen Yao, Chunyu Liu, Liming Liang, Tianxiao Huan, Allan F. McRae, Serkalem Demissie, Sonia Shah, John M. Starr, L. Adrienne Cupples, Panos Deloukas, Timothy D. Spector, Johan Sundström, Ronald M. Krauss, Donna K. Arnett, Ian J. Deary, Lars Lind, Daniel Levy and Erik Ingelsson

*Circ Cardiovasc Genet.* 2017;10:

doi: 10.1161/CIRCGENETICS.116.001487

*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Copyright © 2017 American Heart Association, Inc. All rights reserved.

Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circgenetics.ahajournals.org/content/10/1/e001487>

Free via Open Access

Data Supplement (unedited) at:

<http://circgenetics.ahajournals.org/content/suppl/2017/02/17/CIRCGENETICS.116.001487.DC2>

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Genetics* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

**Reprints:** Information about reprints can be found online at:  
<http://www.lww.com/reprints>

**Subscriptions:** Information about subscribing to *Circulation: Cardiovascular Genetics* is online at:  
<http://circgenetics.ahajournals.org/subscriptions/>

## SUPPLEMENTAL MATERIAL

### *Supplementary Figures*

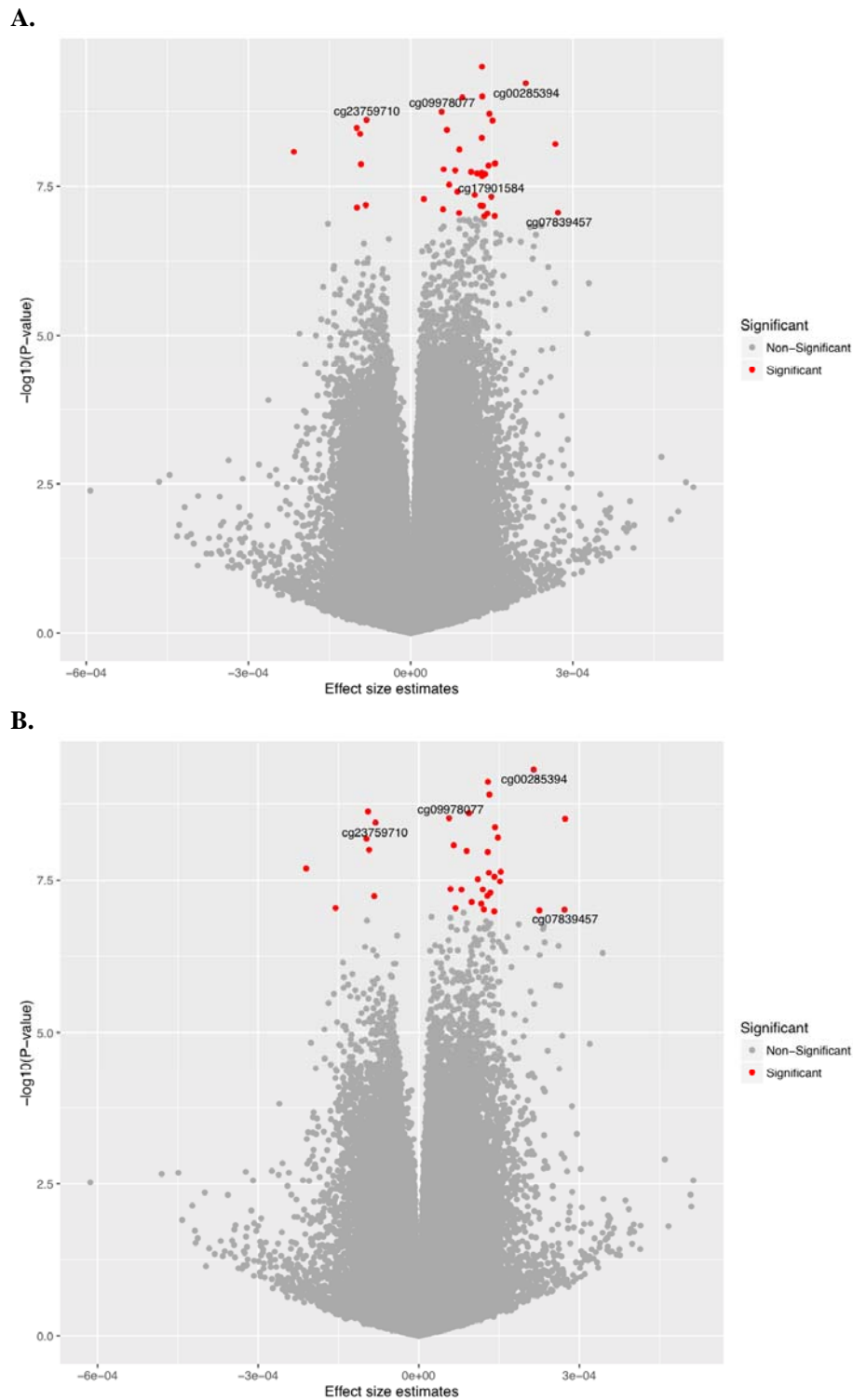
<b>Supplementary Fig. 1.</b> Volcano plots of TC model 1 and 2 in the discovery	<i>Page 4</i>
<b>Supplementary Fig. 2.</b> Volcano plots of LDL-C model 1 and 2 in the discovery	<i>Page 5</i>
<b>Supplementary Fig. 3.</b> Volcano plots of HDL-C model 1 and 2 in the discovery	<i>Page 6</i>
<b>Supplementary Fig. 4.</b> Volcano plots of TG model 1 and 2 in the discovery	<i>Page 7</i>
<b>Supplementary Fig. 5.</b> Manhattan plot of genome-wide analysis of TC (model 1) in the discovery	<i>Page 8</i>
<b>Supplementary Fig. 6.</b> Manhattan plot of genome-wide analysis of LDL-C (model 1) in the discovery	<i>Page 9</i>
<b>Supplementary Fig. 7.</b> Manhattan plot of genome-wide analysis of HDL-C (model 1) in the discovery	<i>Page 10</i>
<b>Supplementary Fig. 8.</b> Manhattan plot of genome-wide analysis of TG (model 1) in the discovery	<i>Page 11</i>
<b>Supplementary Fig. 9.</b> Overlap of DNA methylation sites associated with the blood lipid levels in the discovery.	<i>Page 12</i>
<b>Supplementary Fig. 10.</b> Comparison of effect size estimates between the discovery and replication	<i>Page 13</i>
<b>Supplementary Fig. 11.</b> Comparison of effect size estimates for TC model 1 between the discovery and each of the individual replication cohorts	<i>Page 14</i>
<b>Supplementary Fig. 12.</b> Comparison of effect size estimates for LDL-C model 1 between the discovery and each of the individual replication cohorts	<i>Page 15</i>
<b>Supplementary Fig. 13.</b> Comparison of effect size estimates for HDL-C model 1 between the discovery and each of the individual replication cohorts	<i>Page 16</i>
<b>Supplementary Fig. 14.</b> Comparison of effect size estimates for TG model 1 between the discovery and each of the individual replication cohorts	<i>Page 17</i>
<b>Supplementary Fig. 15.</b> Location of lipid-associated CpGs in relation to CpG islands and genes.	<i>Page 18</i>
<b>Supplementary Fig. 16.</b> Enrichment of lipid-associated CpGs in DHS hotspots from specific tissues using eFORGE tool	<i>Page 19</i>
<b>Supplementary Fig. 17.</b> Overview of significant associations between lipid-associated CpGs and serum metabolites.	<i>Page 20</i>

### *Supplementary Tables*

<b>Supplementary Table 1.</b> Cohort characteristics	<i>Page 21</i>
<b>Supplementary Table 2.</b> CpG sites associated with TC in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate xls</i>
<b>Supplementary Table 3.</b> CpG sites associated with LDL-C in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate xls</i>
<b>Supplementary Table 4.</b> CpG sites associated with HDL-C in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate xls</i>
<b>Supplementary Table 5.</b> CpG sites associated with TG in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate xls</i>
<b>Supplementary Table 6.</b> CpG sites associated with BMI-adjusted TC in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate xls</i>
<b>Supplementary Table 7.</b> CpG sites associated with BMI-adjusted LDL-C	<i>see separate</i>

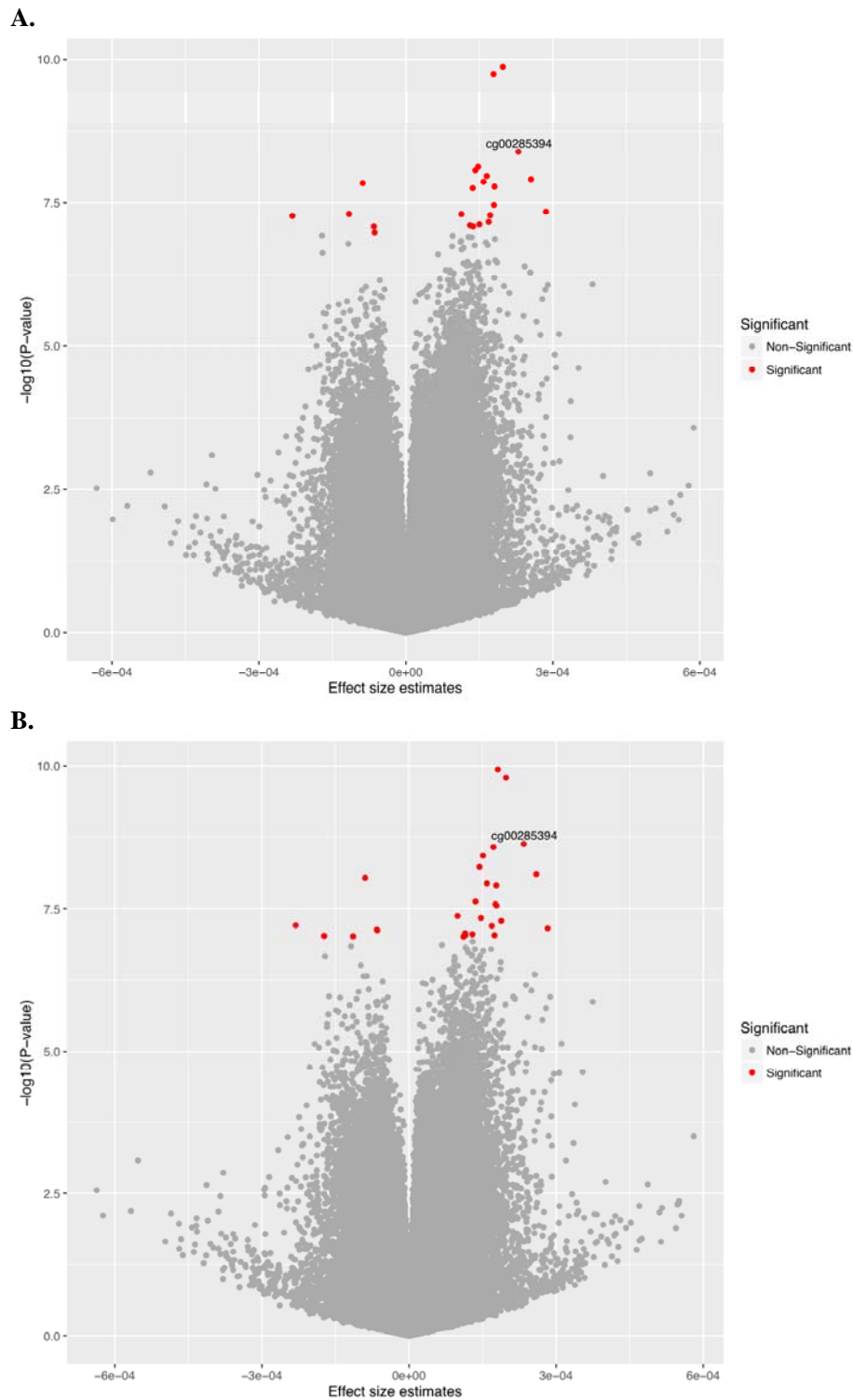


in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>xls</i>
<b>Supplementary Table 8.</b> CpG sites associated with BMI-adjusted HDL-C in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate see separate xls</i>
<b>Supplementary Table 9.</b> CpG sites associated with BMI-adjusted TG in discovery ( $P < 1.08E-07$ ). Sites replicating at Bonferroni-corrected alpha threshold 0.05 are highlighted in bold.	<i>see separate xls</i>
<b>Supplementary Table 10.</b> Association of replicated lipid-associated CpG sites with the same or other traits in published data	<i>see separate xls</i>
<b>Supplementary Table 11.</b> CpG sites associated with TC, LDL-C, HDL-C or TG in combined meta-analysis at Bonferroni-corrected alpha threshold 0.05 ( $P < 1.08E-07$ ).	<i>see separate xls</i>
<b>Supplementary Table 12.</b> CpG sites associated with TC, LDL-C, HDL-C or TG in combined meta-analysis at Bonferroni-corrected alpha threshold 0.05 ( $P < 1.08E-07$ ). Results in table from the secondary BMI-adjusted model.	<i>see separate xls</i>
<b>Supplementary Table 13.</b> Genes annotated to lipid CpGs in GWAS loci of cardiovascular traits	<i>Page 22</i>
<b>Supplementary Table 14.</b> Enriched biological categories among genes annotated to TC-associated CpGs (adjusted P-value $< 0.1$ )	<i>Page 23</i>
<b>Supplementary Table 15.</b> Enriched biological categories among genes annotated to HDL-C associated CpGs (adjusted P-value $< 0.1$ )	<i>Page 24</i>
<b>Supplementary Table 16.</b> Enriched biological categories among genes annotated to TG-associated CpGs (adjusted P-value $< 0.1$ )	<i>Page 25</i>
<b>Supplementary Table 17.</b> Lipid-associated CpGs regulated by genetic sequence variants in <i>cis</i> ( $P < 1E-04$ ) in FHS and replication in PIVUS.	<i>see separate xls</i>
<b>Supplementary Table 18.</b> Lipid-associated CpGs significantly associated with levels of expression of neighbouring genes in blood in FHS (FDR $< 0.05$ ).	<i>Page 26</i>
<b>Supplementary Table 19.</b> Association of lipid-associated CpG sites significant in discovery with blood metabolites (FDR $< 0.01$ ).	<i>see separate xls</i>
<b>Supplementary Table 20.</b> Association of lipid-associated CpGs with incident CHD up to ten years after baseline in FHS and PIVUS.	<i>Page 27</i>
 <b>Supplementary Methods</b>	
Cohort descriptions and sample collections	<i>Page 28</i>
Genome-wide DNA methylation profiling	<i>Page 30</i>
Association of DNA methylation levels in blood cell derived data with lipids	<i>Page 31</i>
Genotyping and Imputation	<i>Page 33</i>
Gene expression profiling	<i>Page 33</i>
Metabolomic profiling	<i>Page 34</i>
 <b>Supplementary References</b>	 <i>Page 34</i>



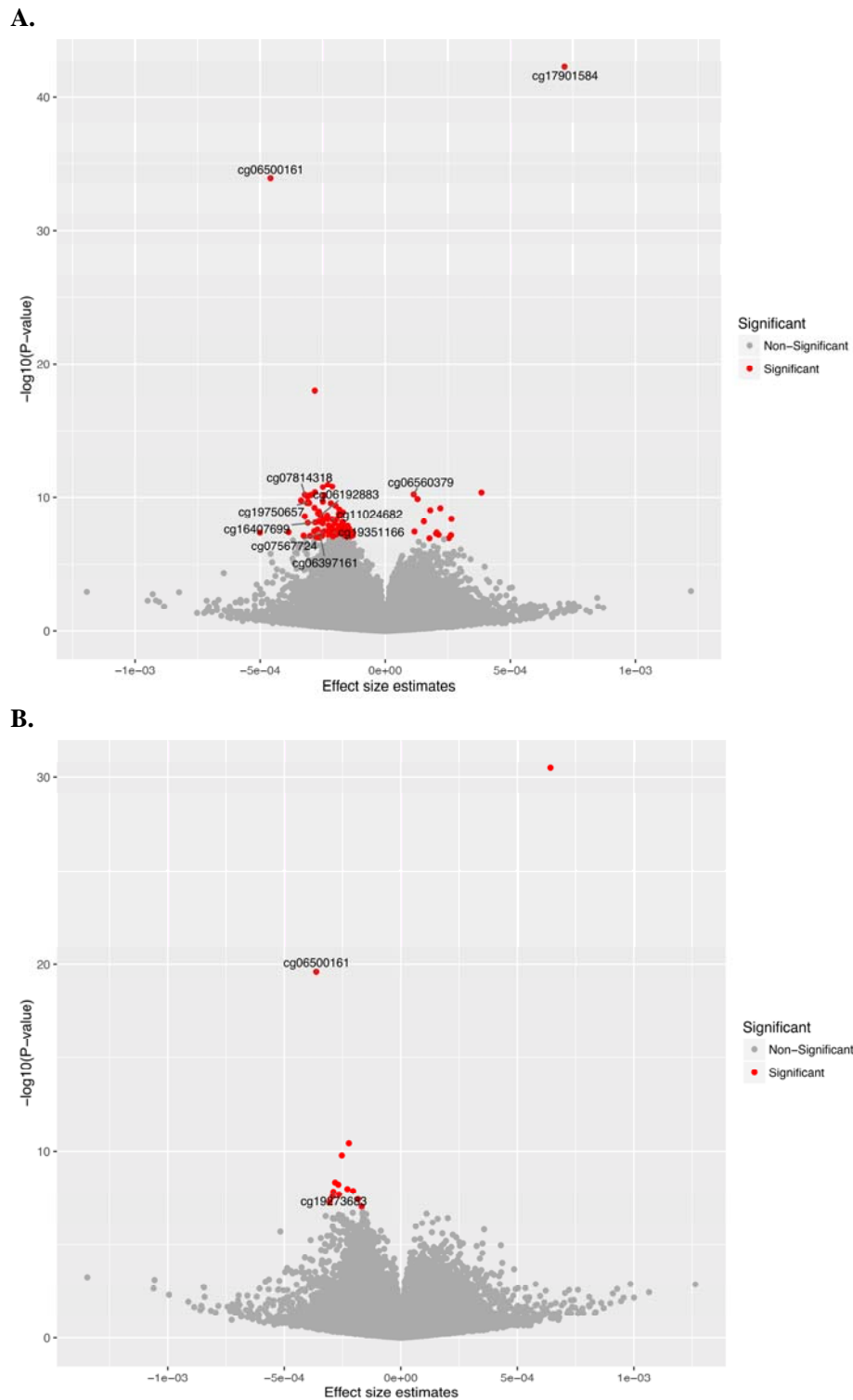
**Supplementary Fig. 1. Volcano plots of TC models 1 and 2 in the discovery.**

**A.** Results of meta-analysis of FHS and PIVUS (discovery) TC model 1. **B.** Results of meta-analysis of FHS and PIVUS (discovery) BMI-adjusted TC model 2. Significant associations are coloured in red. Replicating associations are labelled by CpG marker name in the plot.



**Supplementary Fig. 2. Volcano plots of LDL-C models 1 and 2 in the discovery.**

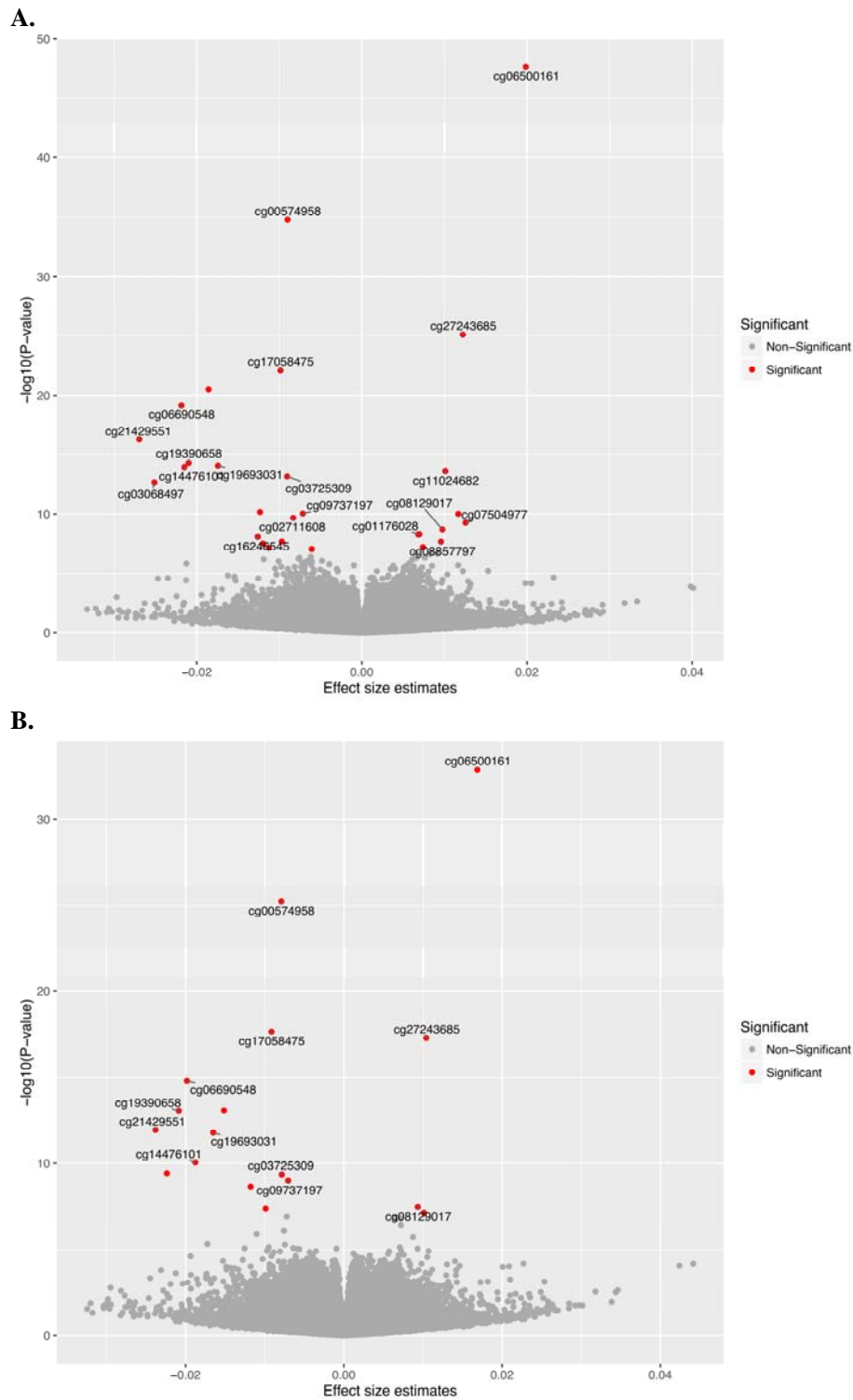
**A.** Results of meta-analysis of FHS and PIVUS (discovery) LDL-C model 1. **B.** Results of meta-analysis of FHS and PIVUS (discovery) BMI-adjusted LDL-C model 2. Significant associations are coloured in red. Replicating associations are labelled by CpG marker name in the plot.



**Supplementary Fig. 3. Volcano plots of HDL-C models 1 and 2 in the discovery.**

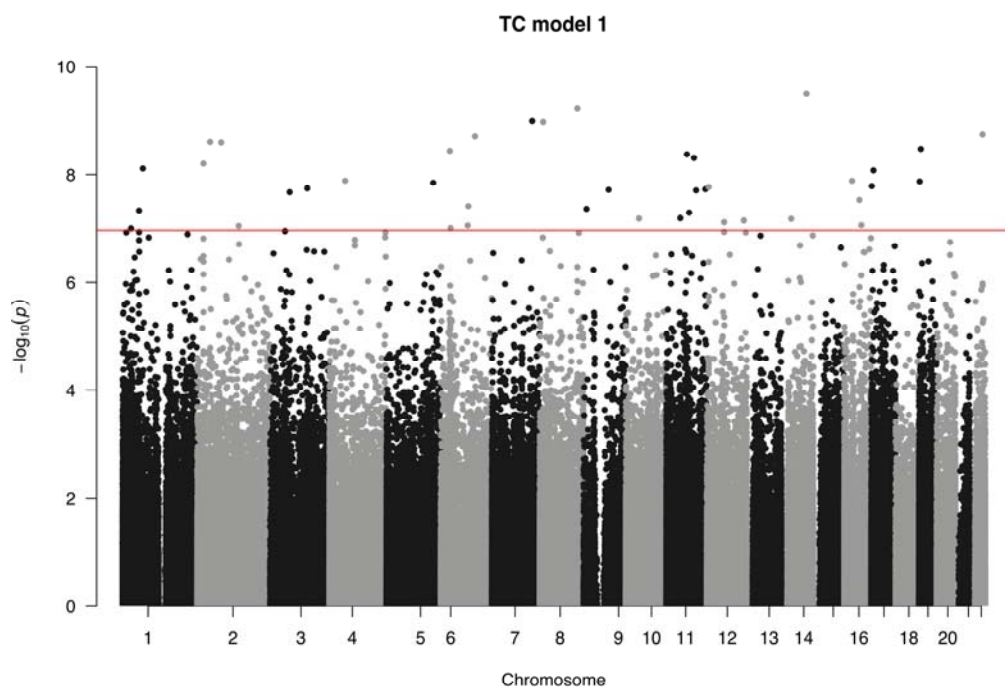
**A.** Results of meta-analysis of FHS and PIVUS (discovery) HDL-C model 1. **B.** Results of meta-analysis of FHS and PIVUS (discovery) BMI-adjusted HDL-C model 2. Significant associations are coloured in red. Replicating associations are labelled by CpG marker name in the plot.



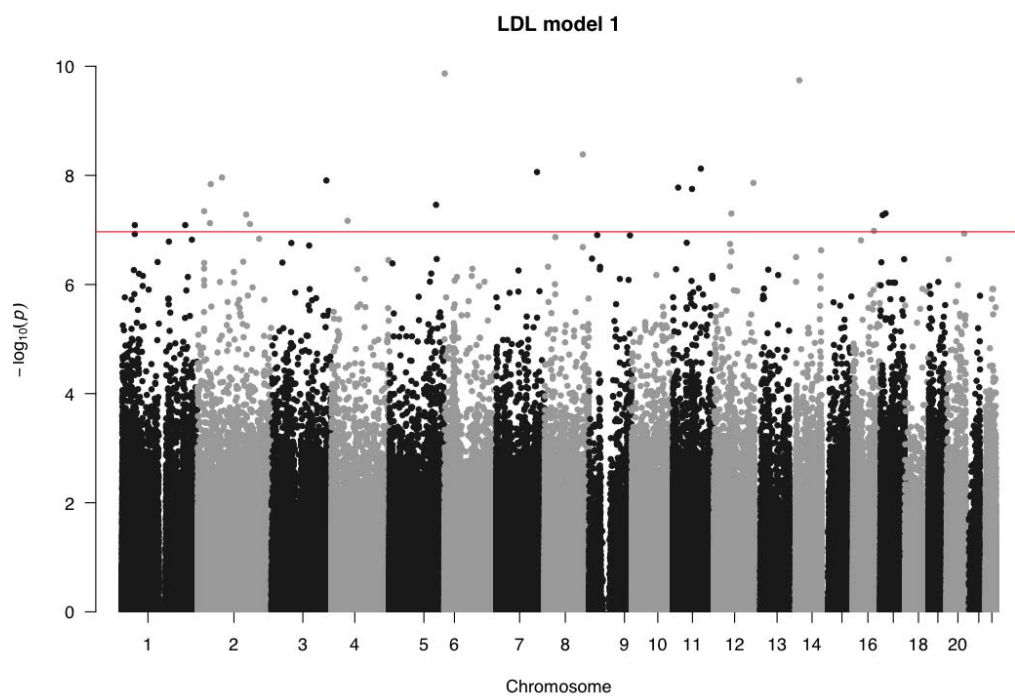


**Supplementary Fig. 4. Volcano plots of TG models 1 and 2 in the discovery.**

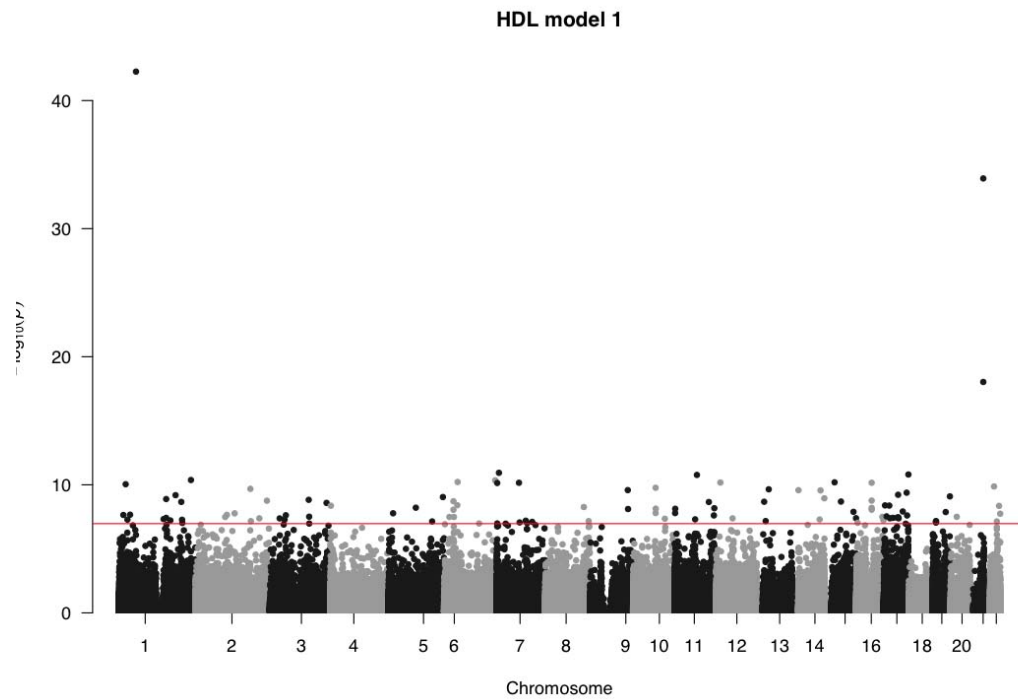
**A.** Results of meta-analysis of FHS and PIVUS (discovery) TG model 1. **B.** Results of meta-analysis of FHS and PIVUS (discovery) BMI-adjusted TG model 2. Significant associations are coloured in red. Replicating associations are labelled by CpG marker name in the plot.



**Supplementary Fig. 5. Manhattan plot of genome-wide analysis of TC (model 1) in the discovery.** The red line indicates the methylome-wide significance level ( $P < 1.08 \times 10^{-7}$ ).

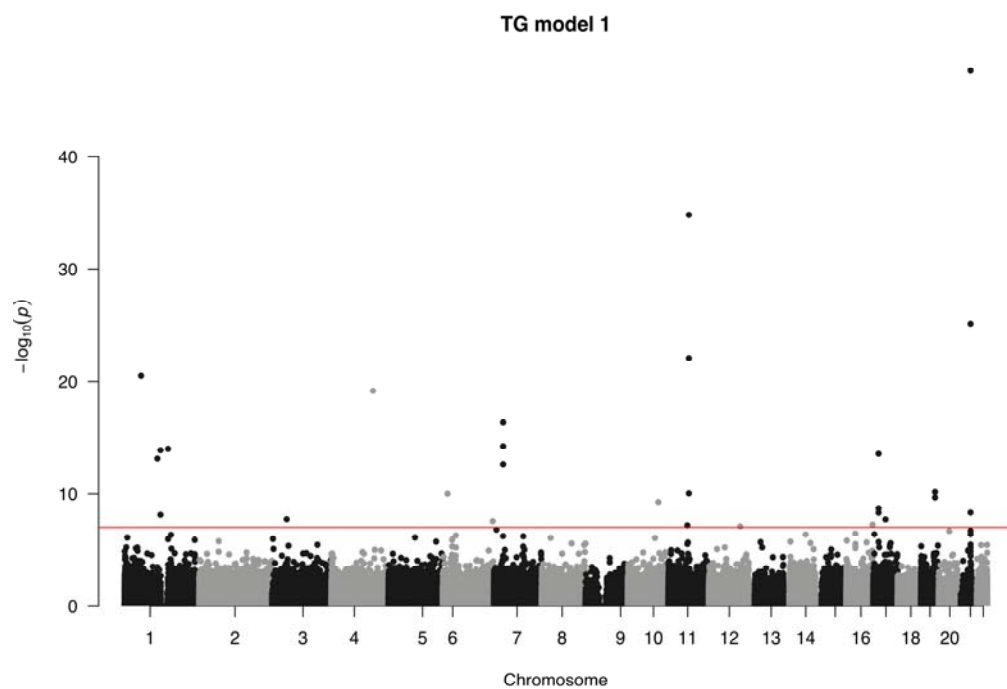


**Supplementary Fig. 6. Manhattan plot of genome-wide analysis of LDL-C (model 1) in the discovery.** The red line indicates the methylome-wide significance level ( $P < 1.08E-07$ ).

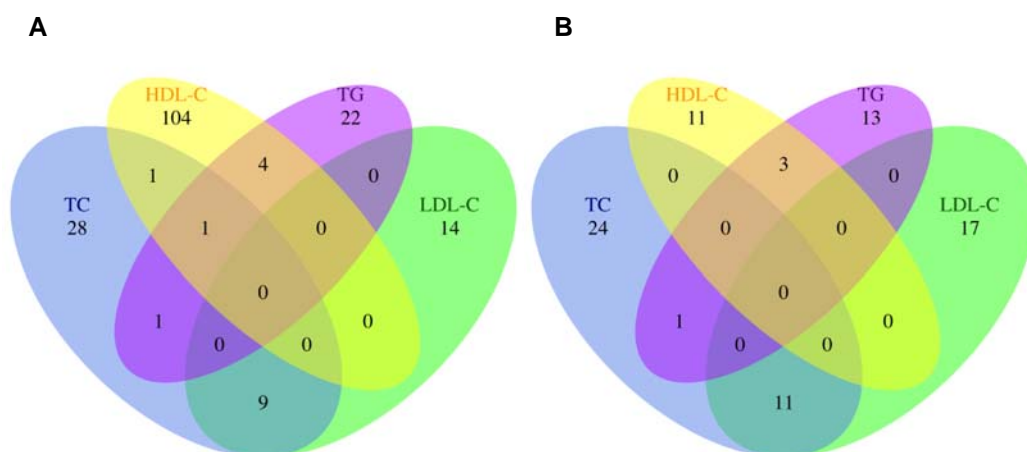


**Supplementary Fig. 7. Manhattan plot of genome-wide analysis of HDL-C (model 1) in the discovery.** The blue line indicates the methylome-wide significance level ( $P < 108E-07$ ).

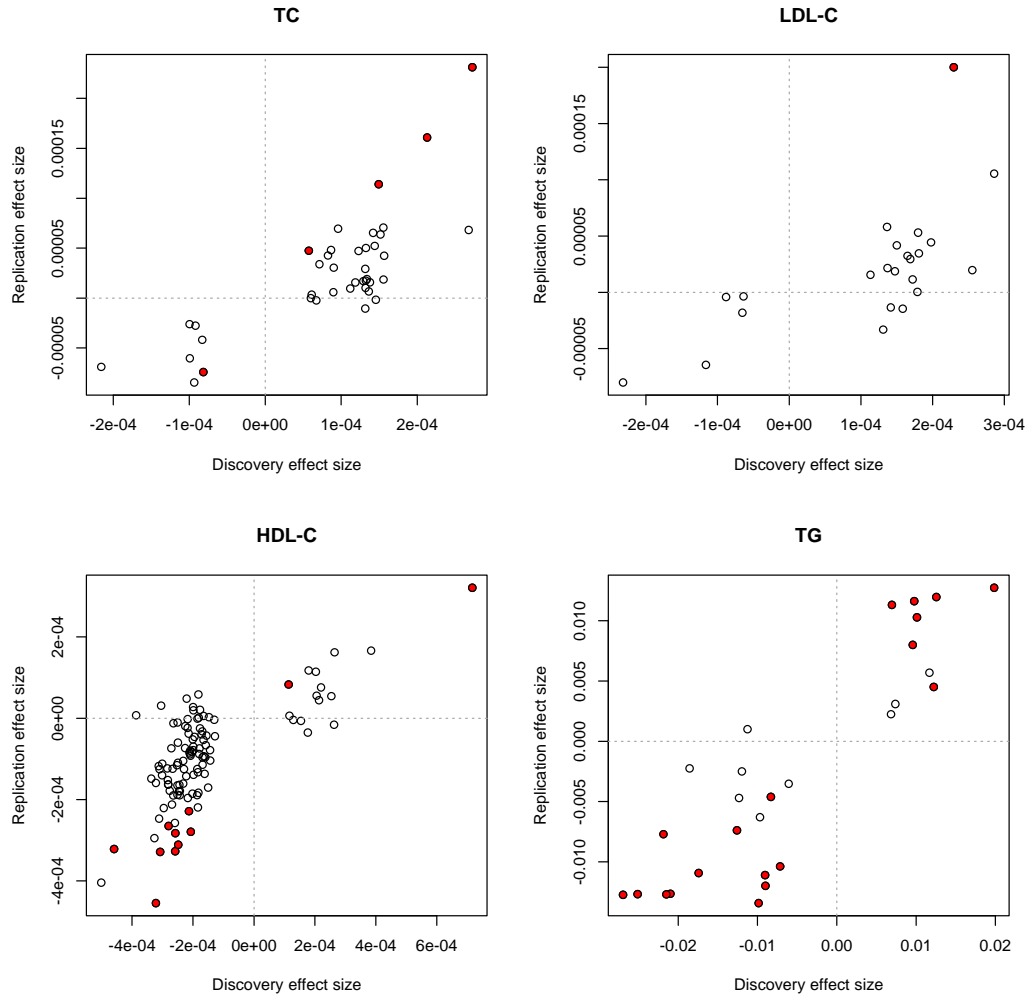




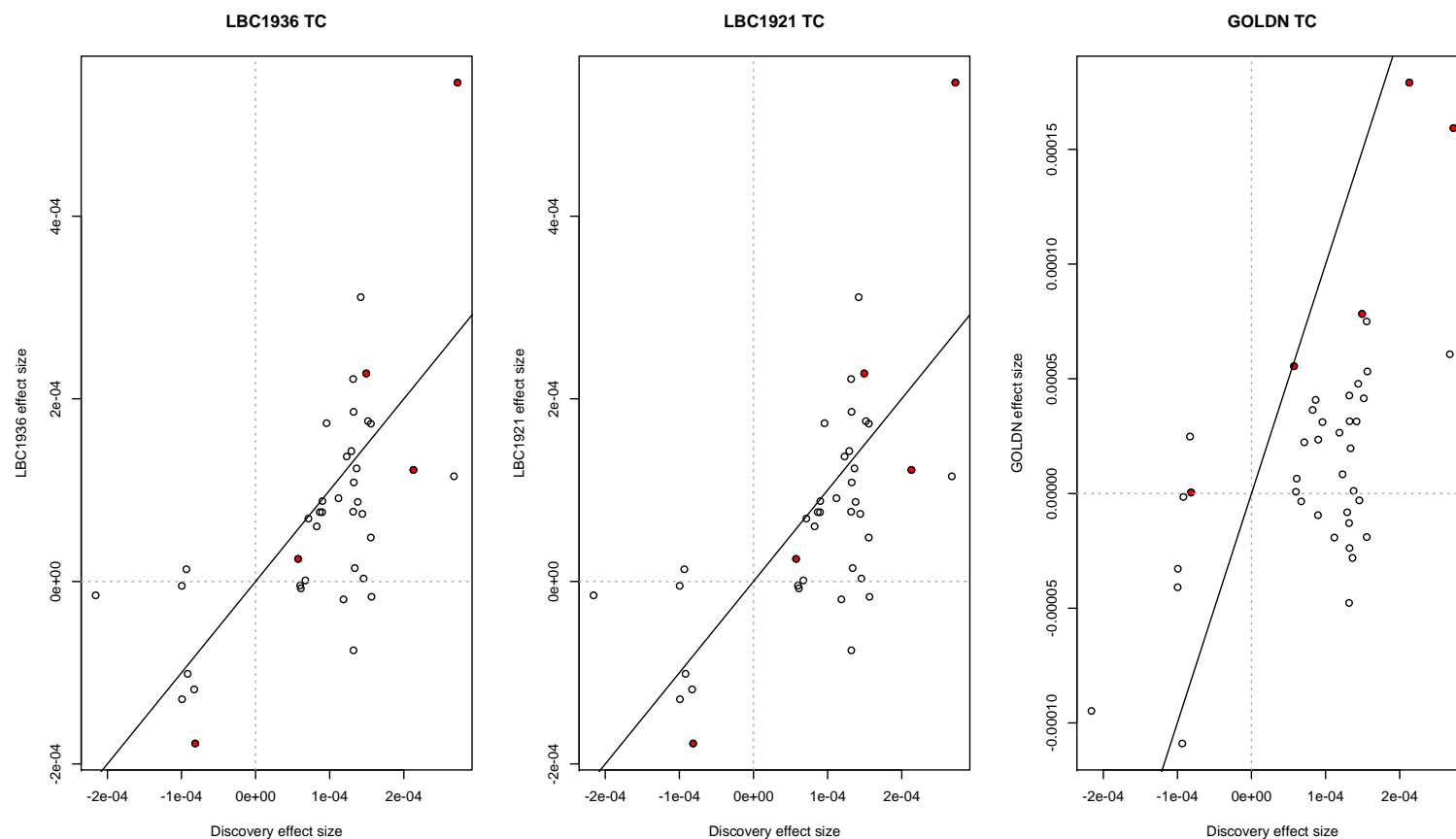
**Supplementary Fig. 8. Manhattan plot of genome-wide analysis of TG (model 1) in the discovery.** The red line indicates the methylome-wide significance level ( $P < 1.08 \times 10^{-7}$ ).



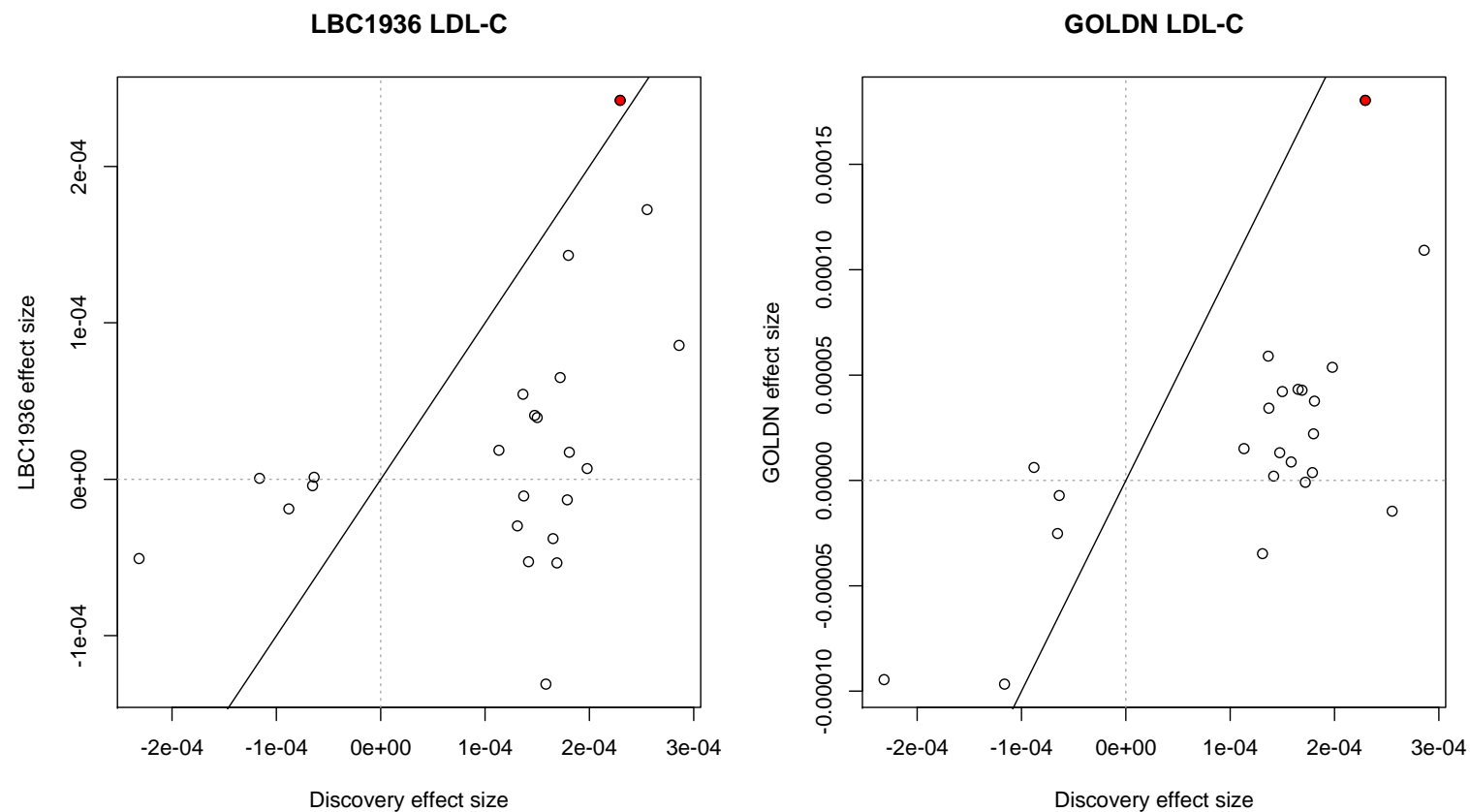
**Supplementary Fig. 9. Overlap of DNA methylation sites associated with the blood lipid levels in genome-wide analyses in the discovery. A: Primary model, B: Secondary BMI-adjusted model.**



**Supplementary Fig. 10. Comparison of effect size estimates between the discovery and replication.** Plot of discovery stage coefficients versus replication stage coefficients for all CpGs significantly associated (per trait) in models 1 in the discovery. **A.** TC, Pearson correlation coefficient ( $r$ ) = 0.78 ( $r$  = 0.81 following removal of replicating sites), **B.** LDL-C,  $r$  = 0.67 ( $r$  = 0.77 following removal of replicating sites), **C.** HDL-C,  $r$  = 0.71 ( $r$  = 0.70 following removal of replicating sites), and **D.** TG,  $r$  = 0.88 ( $r$  = 0.79 following removal of replicating sites). CpG sites significant also in the replication are indicated in red.

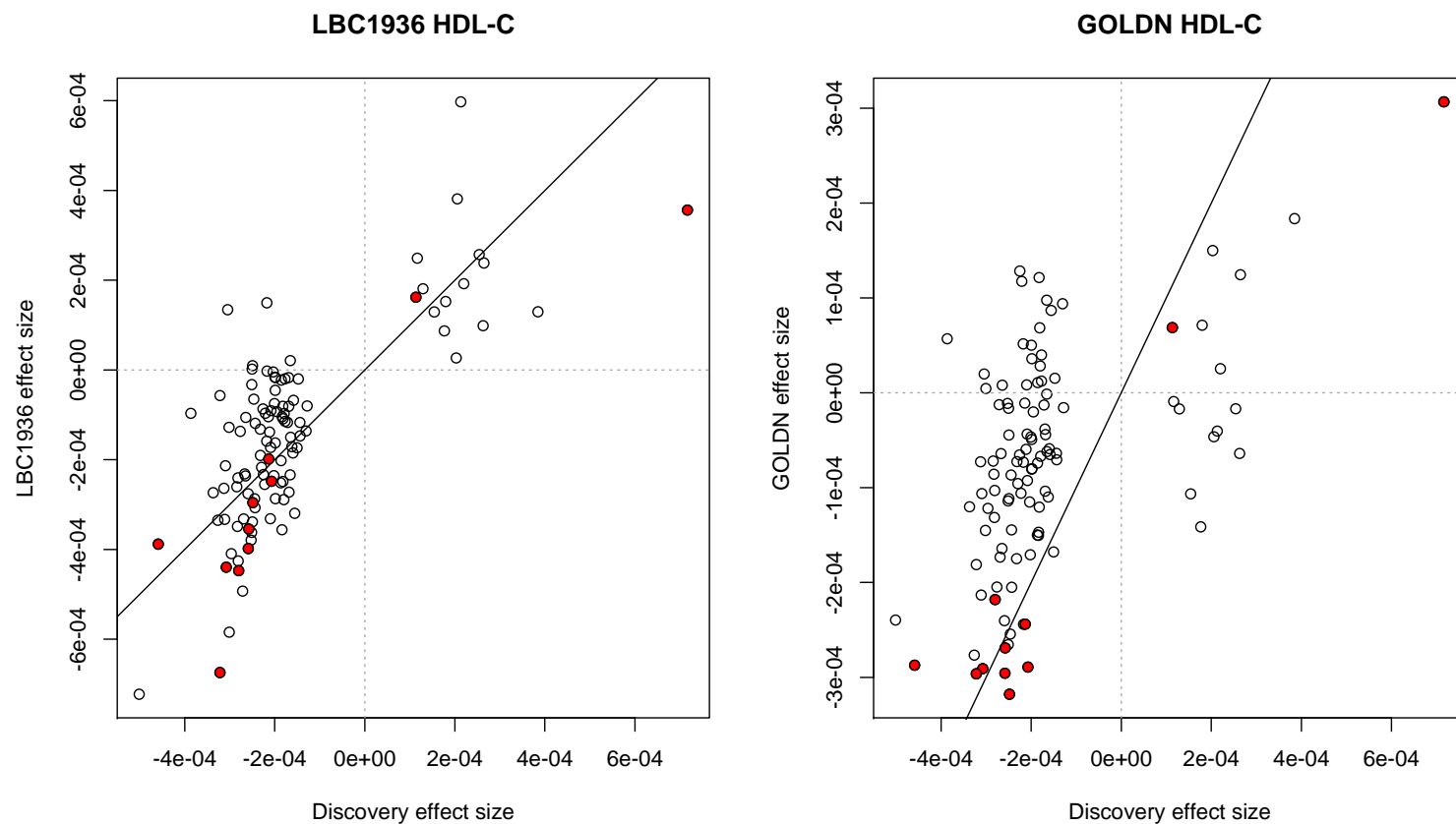


**Supplementary Figure 11. Comparison of effect size estimates for TC model 1 between discovery and each of the individual replication cohorts.** Beta coefficients in figure from lipid EWAS results in Discovery vs. LBC1936 (left panel), Discovery vs. LBC1921 (middle) and Discovery vs. GOLDN (right). CpG sites significant also in the replication are indicated in red. The black line represents the identity line.

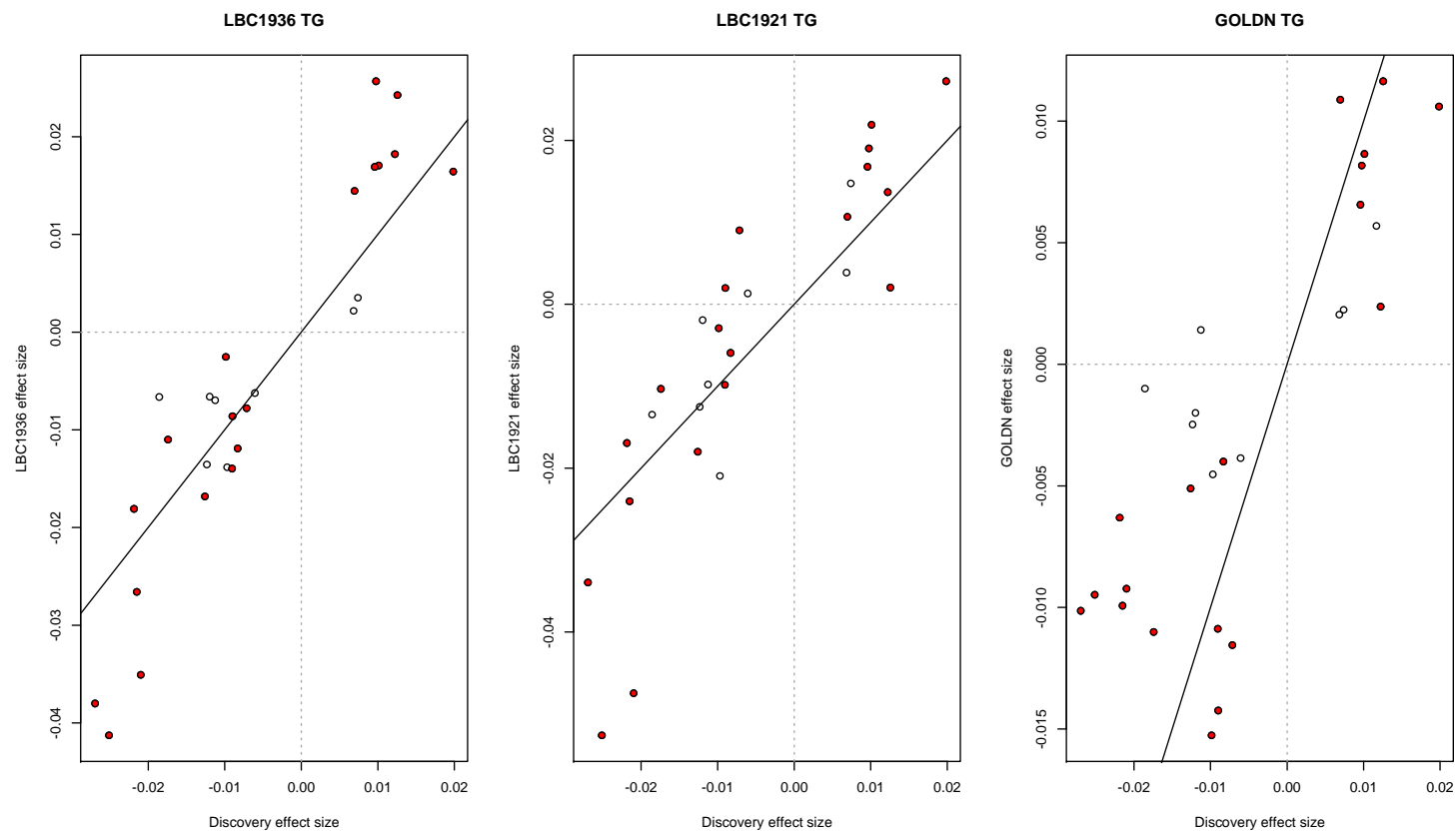


**Supplementary Figure 12. Comparison of effect size estimates for LDL-C model 1 between discovery and each of the individual replication cohorts.** Beta coefficients in figure from lipid EWAS results in Discovery vs. LBC1936 (left panel) and Discovery vs. GOLDN (right). CpG sites significant also in the replication are indicated in red. The black line represents the identity line.

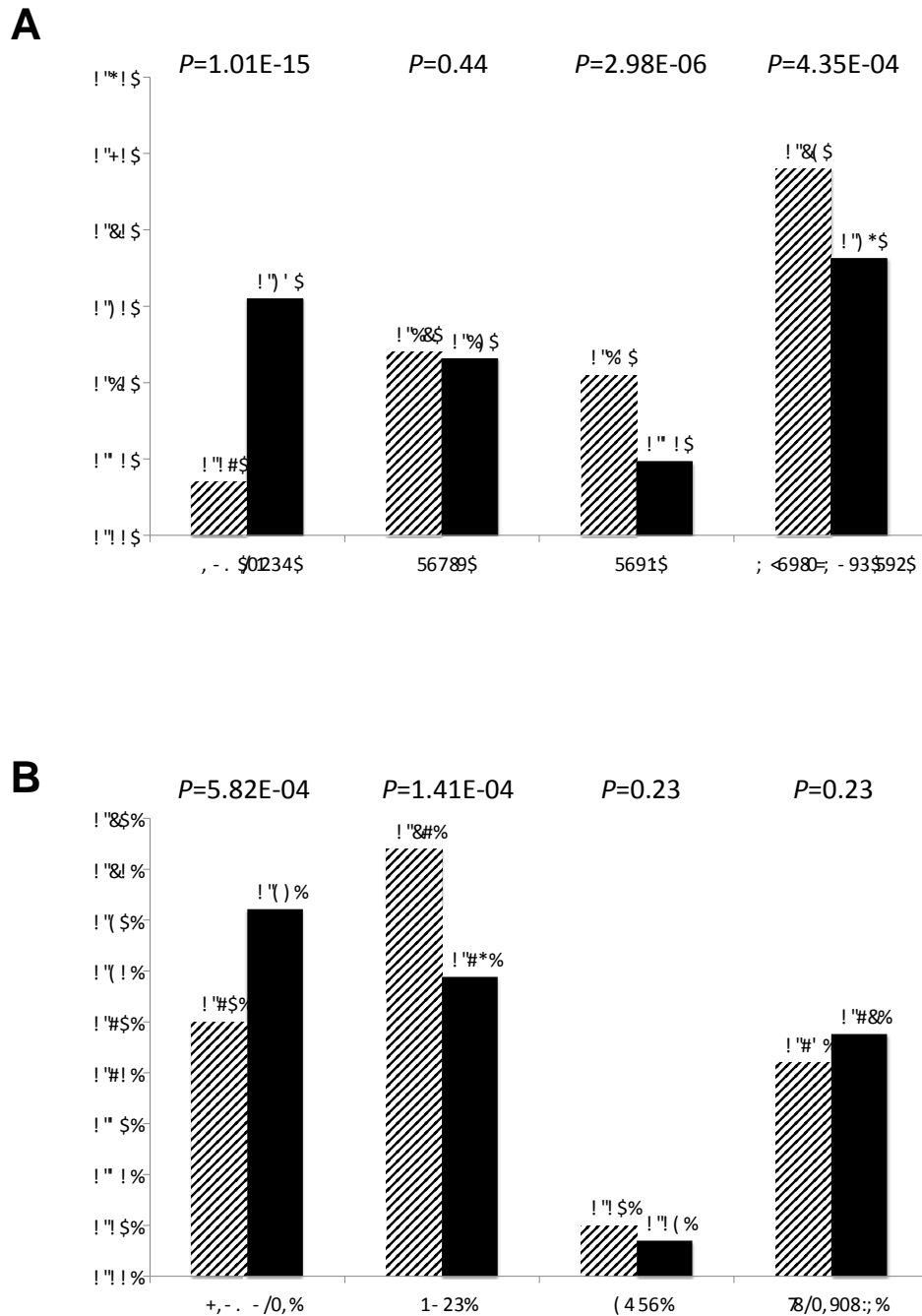




**Supplementary Figure 13. Comparison of effect size estimates for HDL-C model 1 between discovery and each of the individual replication cohorts.** Beta coefficients in figure from lipid EWAS results in Discovery vs. LBC1936 (left panel) and Discovery vs. GOLDN (right). CpG sites significant also in the replication are indicated in red. The black line represents the identity line.



**Supplementary Figure 14. Comparison of effect size estimates for TG model 1 between discovery and each of the individual replication cohorts.** Beta coefficients in figure from lipid EWAS results in Discovery vs. LBC1936 (left panel), Discovery vs. LBC1921 (middle) and Discovery vs. GOLDN (right). CpG sites significant also in the replication are indicated in red. The black line represents the identity line.



**Supplementary Fig. 15. Location of lipid-associated CpGs in relation to CpG islands and genes.** **A.** CpGs were classified into: CpG island, Shore, Shelf and Others/Open sea, and lipid-associated CpGs (striped bars) were compared with all CpGs on the array (black bars). **B.** CpGs were classified into: promoter (TSS1500, TSS200, 5'-UTR, First exon), Body, 3'-UTR and intergenic and lipid-associated CpGs (striped bars) were compared with all CpGs on the array (black bars). Y-axis shows the proportion in each category. P-values in figure are from a one-sided Fisher exact tests.







**Supplementary Table 1.** Cohort characteristics

	<b>FHS</b>	<b>PIVUS</b>	<b>LBC1936</b>	<b>LBC1921</b>	<b>GOLDN</b>
<b>Sample type</b>	Blood	Blood	Blood	Blood	T-cells
<b>Fasting status</b>	Fasted	Fasted	Non-fasted	Non-fasted	Fasted
<b>Age, mean (SD)</b>	66.4 (8.9)	70.2 (0.2)	69.5 (0.8)	79.1 (0.6)	48.8 (16)
<b>BMI (kg/m<sup>2</sup>)</b>	28.3 (5.3)	26.9 (4.3)	27.4 (4.2)	26.1 (4.0)	28.3 (6)
<b>Female (%)</b>	54.3	49.8	52.8	60.3	52.3
<b>TC (mg/dL)</b>					
N	1494	812	654	380	991
Mean (SD)	201 (33)	215.4 (36.9)	224.9 (41.6)	221.7 (43.6)	190.0 (38)
Range	85 - 328	116 - 363.5	123.7 - 417.6	127.6 - 379.0	98- 332
<b>LDL (mg/dL)</b>					
N	1494	810	588	--	991
Mean (SD)	118 (28)	136 (31.9)	135.5 (36.7)	--	121.6 (31)
Range	26-215	42.5-266.8	33.9 - 317.6	--	44 - 236
<b>HDL (mg/dL)</b>					
N	1494	812	592	--	991
Mean (SD)	61 (19)	59.4 (16.7)	60.4 (17.4)	--	47.0 (13)
Range	22-156	23.2-146.9	26.3 - 147.7	--	22- 110
<b>TG (mg/dL)</b>					
N	1494	812	588	376	991
Mean (SD)	112 (65)	109.0 (48.4)	140.8 (64.6)	164.7 (76.9)	137.0 (95)
Range	30-790	11.5-372.0	41.6 - 419.8	62.0 - 611.1	23 - 1085

**Supplementary Table 13.** Genes annotated to lipid CpGs in GWAS loci of cardiovascular traits

Associated lipid trait in epigenetic analysis	Gene annotated to lipid-associated CpG(s)	GWAS trait(s)	Reference
TC, LDL-C	<i>APOB</i>	Lipids (TC, LDL-C), Coronary artery disease	<sup>1, 2</sup>
HDL-C	<i>TRIB1</i>	Lipids (TC, LDL-C, HDL-C, TG), Adiponectin	<sup>1, 3</sup>
HDL-C	<i>LDLR</i>	Lipids (TC, LDL-C), Coronary artery disease	<sup>1, 2</sup>
TG	<i>MYLIP</i>	Lipids (TC, LDL-C)	<sup>1</sup>
LDL-C	<i>GALNT2</i>	Lipids (HDL-C, TG)	<sup>1</sup>
HDL-C	<i>AMPD3</i>	Lipids (HDL-C)	<sup>1</sup>
TC, LDL-C	<i>FADS2</i>	Lipids (TC, LDL-C, HDL-C, TG)	<sup>1</sup>
HDL-C	<i>KLF13</i>	Waist-hip-ratio	<sup>4</sup>
HDL-C	<i>CBX3</i>	Waist-hip-ratio	<sup>4</sup>
TG	<i>LY86</i>	Waist-hip-ratio	<sup>4</sup>
TC	<i>ADCY3</i>	BMI	<sup>5</sup>
TC	<i>PTPRD</i>	Type-2 Diabetes	<sup>6</sup>

**Supplementary Table 14.** Enriched biological categories among genes annotated to TC-associated CpGs (FDR <0.05)

Category	Term	Fold	FDR
<b>Discovery</b>			
GO BP	GO:0016125~ sterol metabolic process	30.44	2.90E-03
GO BP	GO:0008203~ cholesterol metabolic process	33.42	4.00E-03
KEGG Pathway	hsa00100: Steroid biosynthesis	74.78	1.19E-02
GO BP	GO:0008202~ steroid metabolic process	15.22	2.85E-02
Panther BP	BP00019: Lipid, fatty acid and steroid metabolism	5.71	3.66E-02
<b>Replication</b>			
KEGG Pathway	hsa00100: Steroid biosynthesis	299.12	3.34E-03
UP_SEQ_FEATURE	nucleotide phosphate-binding region: FAD	313.33	2.21E-02
REACTOME Pathway	REACT_602: Metabolism of lipids and lipoproteins	22.65	4.41E-02
<b>mRNA expression of gene associated with CpG</b>			
KEGG Pathway	Antigen processing and presentation	30.63	3.76E-02
GO BP	Sterol metabolic process	57.40	4.72E-02

BP, Biological Process

**Supplementary Table 15.** Enriched biological categories among genes annotated to HDL-C associated CpGs (FDR <0.05)

Category	Term	Fold	FDR
<b>Discovery</b>			
N.S.	N.S.	N.S.	N.S.
<b>Replication</b>			
REACTOME Pathway	REACT_602:Metabolism of lipids and lipoproteins	22.65	4.41E-02
<b>mRNA expression of gene associated with CpG</b>			
REACTOME Pathway	REACT_602:Metabolism of lipids and lipoproteins	16.99	5.64E-03

**Supplementary Table 16.** Enriched biological categories among genes annotated to TG-associated CpGs (FDR <0.05)

Category	Term	Fold	FDR
<b>Discovery</b>			
Panther BP	BP00013: Amino acid metabolism	18.110	3.40E-02
<b>Replication</b>			
Panther BP	BP00013: Amino acid metabolism	26.16	5.58E-03
<b>mRNA expression of gene associated with CpG</b>			
Panther BP	BP00013: Amino acid metabolism	26.16	6.09E-03

BP, Biological Process



**Supplementary Table 18.** Lipid-associated CpGs significantly associated with levels of expression of neighbouring genes in blood in FHS (FDR <0.05).

CpG	CGI info	Gene Property	Gene annotated to CpG	Expression of Gene Associated with CpG	Beta	P	meQTL SNP is also an eQTL for associated expressed gene ( $P < 1E-4$ )	meQTL peak SNP	eQTL peak SNP	$r^2$ meQTL SNP - eQTL peak SNP
cg17901584	S_Shore	TSS1500	<i>DHCR24</i>	<i>DHCR24</i>	-1.22	4.84E-13	Yes	rs6687489	rs12131972	0.72
cg19273683		Body	<i>ECE1</i>	<i>ECE1</i>	-0.73	6.75E-08	Yes	rs3026815	rs1067221	0.264
cg01751802	S_Shore	TSS1500	<i>KANK2</i>	<i>KANK2</i>	-0.41	6.50E-05	Yes	19:11317508:TG_T	rs7254270	0.001
cg26313301	S_Shelf	Body	<i>LDLR</i>	<i>LDLR</i>	0.68	8.95E-05	Yes	rs2569550	rs77373181	--
cg14476101	S_Shore	Body	<i>PHGDH</i>	<i>PHGDH</i>	-0.38	2.76E-08	Yes	rs11583993	rs11583993	Identical
cg16246545	S_Shore	Body	<i>PHGDH</i>	<i>PHGDH</i>	-0.41	3.06E-06	Yes	rs11583993	rs11583993	Identical
cg25739016		Body	<i>RCSN1</i>	<i>LOC100128751</i>	2.1	2.34E-09	Yes	rs1229359	rs7513712	0.007
cg22304262	N_Shelf	5UTR/Body	<i>SLC1A5</i>	<i>PRKD2</i>	-0.42	6.91E-07	Yes	rs8105903	rs60652743	0.15
cg08129017	S_Shore	Body	<i>SREBF1</i>	<i>SREBF1</i>	-0.51	1.07E-11	Yes	rs9899634	rs8078756	1
cg09978077	Island	Body	<i>SREBF2</i>	<i>SREBF2</i>	-2.05	9.37E-12	Yes	rs9607850	rs9611674	0.87
cg05295703		--	--	<i>IL1RL1</i>	-3.26	1.42E-18	Yes	rs12469892	rs1420103	0.734
cg05295703		--	--	<i>IL18R1</i>	-1.74	2.86E-06	Yes	rs12469892	rs10490202	0.148
cg24002003		--	--	<i>CHSY1</i>	0.88	9.63E-08	Yes	rs3784526	rs3784526	Identical
cg01176028	N_Shore	Body	<i>ABCG1</i>	<i>ABCG1</i>	-0.66	1.36E-06	No	rs225448	rs9976024	0.17
cg06500161	S_Shore	Body	<i>ABCG1</i>	<i>ABCG1</i>	-1.75	8.58E-49	No	rs225443	rs9976024	0.03
cg27243685	S_Shelf	5UTR/Body	<i>ABCG1</i>	<i>ABCG1</i>	-1.94	7.23E-28	No	rs4148086	rs9976024	0.63
cg22488164	N_Shelf	Body	<i>PLBD1</i>	<i>PLBD1</i>	-0.54	4.84E-05	No	rs2098542	rs151001109	--
cg17501210		Body	<i>RPS6KA2</i>	<i>RNASET2</i>	-0.62	9.75E-10	No	rs7745806	rs429083	>500kb
cg22304262	N_Shelf	5UTR/Body	<i>SLC1A5</i>	<i>SLC1A5</i>	-0.64	3.63E-07	No	rs8105903	rs3027953	0.27
cg11001536		--	--	<i>FAM114A2</i>	-0.62	2.36E-05	N.A. <sup>1</sup>	--	--	--
cg00574958	N_Shore	5UTR	<i>CPT1A</i>	<i>CPT1A</i>	-4.02	5.53E-19	N.A. <sup>1</sup>	--	--	--
cg17058475	N_Shore	5UTR	<i>CPT1A</i>	<i>CPT1A</i>	-2.47	2.20E-10	N.A. <sup>1</sup>	--	--	--
cg09737197	N_Shore	5UTR	<i>CPT1A</i>	<i>CPT1A</i>	-1.58	1.36E-08	N.A. <sup>1</sup>	--	--	--
cg08788930		Body	<i>DENND3</i>	<i>SLC45A4</i>	-1.12	1.20E-05	N.A. <sup>1</sup>	--	--	--
cg21645268	N_Shelf	Body	<i>FDFT1</i>	<i>CTSB</i>	-3.55	1.58E-33	N.A. <sup>1</sup>	--	--	--
cg21645268	N_Shelf	Body	<i>FDFT1</i>	<i>CTSB</i>	1.31	1.57E-06	N.A. <sup>1</sup>	--	--	--
cg18520125		Body	<i>FLT1</i>	<i>FLT1</i>	-0.69	8.94E-07	N.A. <sup>1</sup>	--	--	--
cg16609995		3UTR; TSS1500	<i>GP3M3; PBX2</i>	<i>HLA-DRB6</i>	-3.09	1.92E-06	N.A. <sup>1</sup>	--	--	--
cg16609995		3UTR; TSS1500	<i>GP3M3; PBX2</i>	<i>AGER; RNF5</i>	-0.93	3.12E-06	N.A. <sup>1</sup>	--	--	--
cg16609995		3UTR; TSS1500	<i>GP3M3; PBX2</i>	<i>SLC44A4</i>	-0.67	2.04E-05	N.A. <sup>1</sup>	--	--	--
cg09676013		ncRNA	<i>HLA-DPB2</i>	<i>HLA-DPB1</i>	-1.08	2.77E-13	N.A. <sup>1</sup>	--	--	--
cg09676013		ncRNA	<i>HLA-DPB2</i>	<i>HLA-DPA1</i>	-0.64	5.38E-08	N.A. <sup>1</sup>	--	--	--
cg09676013		ncRNA	<i>HLA-DPB2</i>	<i>HLA-DPB2</i>	0.78	3.81E-05	N.A. <sup>1</sup>	--	--	--
cg03717755		Body	<i>MYLIP</i>	<i>MYLIP</i>	-0.46	1.90E-05	N.A. <sup>1</sup>	--	--	--
cg06690548		Body	<i>SLC7A11</i>	<i>SLC7A11; SLC7A11-AS1</i>	-0.74	3.38E-15	N.A. <sup>1</sup>	--	--	--
cg00285394	S_Shore	Body	<i>SQLE</i>	<i>SQLE</i>	-1.42	8.28E-13	N.A. <sup>1</sup>	--	--	--

1. No significant meQTL

**Supplementary Table 20.** Association of 33 replicating lipid-associated CpGs with incident CHD<sup>1</sup> up to ten years after baseline in FHS and PIVUS.

CpG	Gene	Gene Property	Meta-analysis			FHS		PIVUS		Trait(s) <sup>3</sup>	Direction CpG - lipid
			HR (95% CI) <sup>2</sup>	P	Heterogeneity P	HR (95% CI) <sup>2</sup>	P	HR (95% CI) <sup>2</sup>	P		
cg27243685 <sup>1</sup>	ABCG1	5UTR/Body	1.38 (1.15-1.66)	6.86E-04	0.56	1.41 (1.12-1.77)	3.40E-03	1.33 (0.96-1.84)	0.079	TG (m1&m2)	+

1. 115 events in FHS and 64 events in PIVUS  
2 Hazard ratio per SD increment in methylation at CpG site.  
3. m1: primary model, m2: secondary BMI-adjusted model

## Supplementary Methods

### Cohort descriptions and sample collections

#### *Discovery cohorts*

**FHS Offspring Cohort** was initially recruited in 1971 and included 5,124 offspring (and their spouses) from the FHS Original Cohort <sup>7</sup>. In the FHS, the eligible sample for this investigation was from the 3,021 participants in the FHS Offspring Cohort who attended the eighth examination cycle in 2005-2008. The included sample was determined based on the number of participants consenting to genomic studies with available DNA and methylation assays passing quality control measures. The DNA methylation, anthropometric, and laboratory measures were obtained from the same examination. At each study visit, participants underwent a routine physical examination and a medical history interview. Participants were asked to bring in their current medication containers. Weight was measured to the nearest pound with the participant wearing only a gown without slippers or shoes, standing in the middle of the scale (Detecto Scale, Worchester Scale) with weight equally distributed on both feet. Standing height was measured to the nearest ¼ inch, with the participant barefoot or wearing thin socks, using a vertical mounted stadiometer. BMI was calculated as weight in kg divided by height in meters squared. Peripheral blood samples were collected in the morning from participants after an eight-hour fast. TC, HDL-C and TG were measured via an enzymatic colorimetric assay (Roche Hitachi 911, Roche Diagnostics) and LDL-C was calculated by the Friedewald equation. CHD was defined as a fatal or non-fatal myocardial infarction (MI), coronary death, revascularization procedure (percutaneous transluminal coronary angioplasty or coronary artery bypass graft) or coronary insufficiency (unstable angina). All CHD events were reviewed and adjudicated by a physician endpoint committee.

**PIVUS** is a prospective community-based cohort of participants from Uppsala, Sweden. All men and women at age 70 living in Uppsala in 2001 were invited to participate. The 1,016 participants (50% women) have been extensively phenotyped, as described previously <sup>8</sup>, and on the Internet ([www.medsci.uu.se/pivus/](http://www.medsci.uu.se/pivus/)). The eligible sample for investigation was from the 1,016 enrolled patients at 70 years of age and conducted between the years of 2001-2003. Lipid traits were measured at Uppsala University Hospital using routine medical chemistry methods. LDL-C was calculated by the Friedewald equation. The participants have been re-examined at ages 75 and 80, and their morbidity and mortality has been followed via national registers and journal review. Clinical diagnoses by journal review of CVD up to 10 years after baseline were used to define disease events. In the present study, we combined acute fatal or non-fatal MI and revascularization procedure (percutaneous transluminal coronary angioplasty or coronary artery bypass graft) into a composite atherosclerotic CHD endpoint.

#### *Replication cohorts*

**The LBC 1921 and 1936** are two longitudinal studies of ageing <sup>9-11</sup>. They derive from the Scottish Mental Surveys of 1932 and 1947, respectively, when nearly all 11-year old children in Scotland completed a test of general cognitive ability <sup>9</sup>. Survivors living in the Lothian area of Scotland were recruited in late-life at mean age 79 for LBC1921 (n=550) and mean age 70 for LBC1936 (n=1,091). Follow-up has taken place at ages 70, 73, and 76 in LBC1936 and ages 79, 83, 87, and 90 in LBC1921. The eligible sample for investigation was collected between at age 70 for LBC1936 and age 79 for LBC1921. Collected data include genetic information, longitudinal epigenetic information, longitudinal brain imaging (LBC1936), and numerous blood biomarkers, anthropomorphic and lifestyle measures. Serum cholesterol was measured as part of a blood analysis profile. Non-fasting blood was drawn on the day of cognitive assessment and analysed within 24 h in serum stored at 4 °C using an enzymatic Quinoneimine dye method measuring at 500 nm, at the Western General Hospital, Edinburgh. For LBC1921 HDL-C and LDL-C were not available.

**GOLDN:** The National Heart, Lung, and Blood Institute GOLDN study was designed to identify genetic determinants of lipid response to two interventions (a high-fat meal challenge and fenofibrate treatment for 3 weeks). The GOLDN study has been previously described in detail in Irvin *et al* <sup>12</sup>. Briefly, the study ascertained and recruited families from the Family Heart Study at two centres, Minneapolis, MN and Salt Lake City, UT, who self-reported to be white. Only families with at least two siblings were recruited for a total of 1,327 individuals. Volunteers were required to withhold lipid-lowering agents (pharmaceuticals or nutraceuticals) for at least 4 weeks prior to the initial visit to be eligible. A total of 1,053 met all eligibility requirements. The study protocol was approved by Institutional Review Boards at the University of Minnesota, University of Utah, and Tufts University/New England Medical Center. For the current study, we evaluated fasting blood lipids among 991 participants for whom baseline epigenetic data were available. Lipids were measured before the diet and drug intervention. Participants were asked to fast for  $\geq 12$  hours and abstain from alcohol intake for  $\geq 24$  hours. TG was measured by a glycerol-blanked enzymatic method (Trig/GB, Roche Diagnostics Corporation, Indianapolis, IN). TC was measured using a cholesterol esterase–cholesterol oxidase reaction (Chol R1, Roche Diagnostics Corporation) on the Roche/Hitachi 911 Automatic Analyzer (Roche Diagnostics Corporation). The same reaction was also used to measure HDL-C after precipitation of non-HDL-C with magnesium/dextran. LDL-C was measured by a homogeneous direct method (LDL Direct Liquid Select™ Cholesterol Reagent, Equal Diagnostics, Exton, PA). Data on medical history, physical activity and other lifestyle factors such as alcohol intake, smoking status, and diet were collected using an interviewer-administered questionnaire. Weight was measured by a beam balance and height was ascertained by a stadiometer. BMI was calculated as weight in kilograms divided by height in meters squared.

#### **Genome-wide DNA methylation profiling**

**FHS:** Buffy coat preparations were obtained from the whole blood samples and genomic DNA was extracted using the Gentra Puregene DNA extraction kit (Qiagen, Venlo, Netherlands). DNA samples were bisulphite converted using the EZ DNA Methylation kit (Zymo Research, Irvine, CA) and analysed on Illumina HumanMethylation450 chips (Illumina Inc., San Diego, CA, USA) following the manufactures' protocol. DNA methylation arrays were run in two laboratory batches at the John's Hopkins Center for Inherited Disease Research (lab batch #1) and University of Minnesota Biomedical Genomics Center (lab batch #2). The first batch included 576 samples from an earlier CVD case-control study <sup>13</sup> and the second batch included 2,270 samples from the remainder of the Offspring cohort participants. DNA methylation data were normalised within laboratory batches using the DASEN methodology implemented in the *wateRmelon* package <sup>14</sup> in R (version 3.0.2), which includes background adjustment of the methylated and unmethylated intensities and quantile normalisation of the methylated and unmethylated probes within the two types of probe technologies separately. Samples with a missing rate  $>1\%$  at  $P<0.01$  ( $n=10$  for batch #1 and  $n=35$  for batch #2), poor single nucleotide polymorphism (SNP) matching to the 65 SNP control probe locations ( $n=38$  for batch #1 and  $n=41$  for batch #2), and outliers by multi-dimensional scaling techniques ( $n=25$  for batch #1 and  $n=48$  for batch #2) were excluded. Probes with missing rate  $>20\%$  at  $P<0.01$  ( $n=466$  from batch #1 and  $n=366$  from batch #2), as well as probes previously identified to map to multiple locations <sup>15</sup> or to have an underlying SNP (minor allele frequency [MAF]  $>5\%$  in European ancestry (EUR) 1000 genomes project data) at the CpG site or within 10 bp of the single base extension ( $n=42,251$ ) were excluded. Following quality control, DNA methylation data from 2,377 FHS participants and 443,252 probes remained for analyses. The FHS methylation data are available at dbGaP under the accession number phs000724.v2.p9.

**PIVUS:** Blood for DNA methylation assay were collected at baseline (at age 70). Genomic DNA was extracted from blood samples and bisulphite conversion of 500ng genomic DNA was performed using the EZ-96 DNA Methylation Gold Kit (Zymo Research Product). The equivalent of approximately 200ng of bisulphite converted DNA, was removed, evaporated to a volume of  $<4\mu\text{l}$ , and used for methylation profiling using the Illumina Infinium assay and

the Illumina HumanMethylation450 v.1.2 bead chip according to the protocol from the supplier (Illumina Inc., San Diego, CA, USA). The results were analysed with GenomeStudio 2011.1 (Illumina Inc., San Diego, CA, USA). After exclusion of replicates a total of 1,002 study participants had methylation data available for quality control procedures. Three samples were excluded based on poor bisulphite conversion efficiency, twelve samples due to low pass rate of CpG sites (<98.5% with a detection  $P>0.01$ ) and a further six samples based on low SNP genotype match (>1 SNP mismatches) between genotypes from the methylation array and Omni/MetaboChip genotyping chips leaving 981 individuals with adequate methylation data available for analyses. Following removal of participants with abnormal leukocyte cell counts (>10x10<sup>9</sup> cells/L; n=14) methylation data from 967 individuals remained for analyses. The signal intensities for the methylated and unmethylated state were then quantile normalised for each probe type separately, and beta values were calculated.

**LBC 1921 and 1936:** Detailed information about the collection and QC steps undertaken on the LBC methylation data has been reported previously<sup>16</sup>. Briefly, the Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA) was used to measure DNA methylation in whole blood of consenting participants. Background correction was performed using the R *minfi* package<sup>17</sup> and QC was used to remove probes with a low detection rate (<95% detection rate at  $P<0.01$ ), probes with low quality (manual inspection), samples with a low call rate (samples with <450,000 probes detected at  $P<0.01$ ), and samples with a poor match between genotypes and SNP control probes (cross-checked using *wateRmelon* package<sup>14</sup>) or incorrect predicted sex. Post QC, DNA methylation data were available for 446 LBC1921 participants at age 79 and for 920 LBC1936 participants at age. The LBC methylation data are available at European Genome-Phenome Archive under accession number EGAS00001000910.

**GOLDN:** DNA was extracted from CD4<sup>+</sup> T-cells harvested from stored buffy coats using antibody-linked Invitrogen Dynabeads<sup>18</sup>. Stored buffy coats were collected at the same time lipid concentrations were measured. We lysed cells captured on the beads and extracted DNA using DNeasy kits (Qiagen, Venlo, Netherlands) and methylation was assayed across ~470,000 autosomal CpG sites using the Illumina Infinium Human Methylation450 Beadchip (Illumina, San Diego, CA). For each assay, 500ng of DNA was treated with sodium bisulfite (EZ DNA, Zymo Research, Irvine, CA) prior to standard Illumina amplification, hybridization, and imaging steps. The resulting intensity files were analyzed with Illumina's GenomeStudio which generated beta scores (i.e. the proportion of total signal from the methylation specific probe or color channel) and "detection  $P$ -values" (the probability that the total intensity for a given probe falls within the background signal intensity). Beta scores with an associated detection  $P$ -value greater than 0.01 were removed and samples with more than 1.5% missing data points were eliminated from further analysis. Furthermore, any CpG probes where more than 10% of samples failed to yield adequate intensity were removed. A total of 58 samples were removed. The filtered beta scores were then subjected to batch normalization with the ComBat package for R software in non-parametric mode<sup>19</sup>. We performed the normalization in parallel on random subsets of 20,000 CpGs per run where each array of 12 samples was used as a "batch." These methods have been extensively described in Absher *et al* and the utility of ComBat to correct for batch effects in comparison to other programs is reported<sup>20, 21</sup>. To correct for probe chemistry, we separately normalized probes from the Infinium I and II chemistries and subsequently adjusted the  $\beta$  scores for Infinium II probes using the equation derived from fitting a second order polynomial to the observed methylation values across all pairs of probes located <50bp apart (within-chemistry correlations >0.99), where one probe was Infinium I and one was Infinium II. Finally, we eliminated any CpGs where the probe sequence mapped either to a location that did not match the annotation file or to more than one locus. We identified such markers by re-aligning all probes (with unconverted Cs) to the human reference genome. After these quality control procedures, there were methylation data from 461,281 CpGs. Principal components (PCs) based on the beta scores of all autosomal CpGs passing QC were generated using the *prcomp* function in R (V 2.12.1) and used to adjust for cell purity in association analysis. Deconvolution estimated CD4<sup>+</sup> T-cell percentages were calculated adapting the method of Abbas *et al*<sup>22</sup>. Predicted CD4<sup>+</sup> T-cell



percent purity was highly correlated with PC1 ( $r^2=0.85$ ,  $P=4E-293$ ) but not other PCs, thus supporting the usefulness of methylation PCs in adjusting for cell purity in our analysis. The GOLDN methylation data are available at dbGaP under the accession number phs000741.v1.p1.

#### **Association of methylation of blood cell-derived DNA with lipid levels**

**FHS:** Linear mixed effects regression models were conducted to test the association between site-specific DNA methylation and each lipid phenotype (TC, HDL-C, LDL-C, TG) individually. Participants currently on lipid lowering medication were excluded. The primary model was adjusted for age, sex and technical covariates. The secondary model was additionally adjusted for BMI. Technical covariates included chip, row, column (specified as random effects) and two methylation principal components (PCs) to adjust for unmeasured batch effects. Cell count heterogeneity was accounted for by adjusting for imputed cell counts obtained by the Houseman method<sup>23</sup>. The linear mixed effect models were run with the *pedigreemm* package in R (version 3.1), which additionally accounts for the family correlation structure in the FHS. Familial relatedness was obtained by reported relationships and genetic similarity calculated by identity-by-descent (IBD) probabilities. For any inconsistencies between reported and IBD relationships, relationships obtained from IBD probabilities were utilized. After removing individuals on lipid-lowering medication and thus with missing phenotype, a total of 1,494 participants were considered in the association analyses.

**PIVUS:** The associations between normalised DNA methylation beta values and phenotypes were modelled by a linear mixed effect model, using R<sup>24</sup> and the *lmer* function (*lme4* package), fitted by maximum-likelihood assuming a normally distributed error term. Models were adjusted for age, sex and predicted white cell counts (estimated from the DNA methylation data using the Houseman algorithm<sup>23</sup> as implemented in R package *minfi* for Illumina HumanMethylation450<sup>17</sup>) as fixed effects and chip, chip row and chip column as random effects. A likelihood ratio test was used to assess the significance of the phenotype effect. The *p-value* of the phenotype effect in each model was calculated from the Chi-square distribution with 1 degree of freedom using  $-2\log(\text{likelihood ratio})$  as the test statistic. After removing individuals on lipid-lowering medication ( $n=155$ ), a total of 812 individuals were considered in the association analyses.

**LBC 1936 and 1921:** Linear regression modelling was used to assess the association between DNA methylation (outcome variable) and the lipid traits (predictor variable). Covariates included age, sex, and measured white blood cell counts (eosinophils, basophils, neutrophils, monocytes, and lymphocytes). Additional adjustments were made for BMI in secondary models. Participants were excluded from the analyses if they were taking lipid-lowering medication. All statistical analyses were performed using R software (<http://cran.r-project.org/>). After removing individuals on lipid-lowering medication and thus with missing phenotype, a total of 654, 588, 592 and 588 participants were considered in the association analyses of TC, LDL-C, HDL-C and TG, respectively for LBC1936. For LBC1921, a total of 380 and 376 were considered in the association analyses of TC and TG, respectively.

**GOLDN:** Associations between normalised methylation beta values at each CpG site and lipid traits were analysed using mixed linear regression models adjusted for age, gender, study site, and 4 methylation PCs (as a proxy for cell purity) as fixed effects and family structure as a random effect using the R kinship package (*lmeKin* function). A second set of models additionally adjusted for BMI. After removing four observations due to missing phenotype or covariate data, a total of 991 participants were considered in the association analysis.

#### **Genotyping and Imputation**

**FHS:** SNP data were obtained from the Affymetrix 550K Array (Affymetrix, Santa Clara, CA) and imputed to 1000 Genomes SNPs (phase 1 release), as previously reported<sup>25</sup>. The FHS genotype data are available at dbGaP under the accession number phs000342.v13.p9.

**PIVUS:** Individuals were genotyped using the Illumina OmniExpress and Illumina MetaboChip microarrays. Prior to imputation, quality control was performed. Exclusion of

samples were performed based on the following criteria: genotype call rate <95%; heterozygosity >3 SD; gender discordance; duplicated samples; identity-by-descent match; and ethnic outliers. Monomorphic SNPs; or SNPs with Hardy-Weinberg equilibrium  $P < 1 \times 10^{-6}$ ; genotype call rate <0.99 (SNPs with MAF <5%) or <0.95 (SNPs with MAF  $\geq$  5%); MAF <1% were excluded from analysis. Data were imputed to 1000G (version: March 2012) using Impute v.2.2.2<sup>26</sup>.

### Gene expression profiling

Gene expression data were available for participants in the FHS. RNA was extracted from whole blood using the PAXgene Blood RNA System Kit (Qiagen, Venlo, Netherlands) with mRNA expression profiling assessed using the Affymetrix Human Exon 1.0 ST GeneChip platform. Gene expression data were normalised using robust multichip average methods<sup>27</sup> with quality control measures as previously described<sup>13</sup>. Cell count proportions were derived from gene expression markers in this sample set as there was overlap between gene expression measures and directly measured cell counts (lymphocytes, monocytes, neutrophils, basophils, and eosinophils) from a sample of 2,280 Third Generation FHS participants obtained during the second examination cycle (2008-2011). Internal validation using training and testing datasets achieved an  $r^2 > 0.8$  in the majority of cell lines (except basophils). The FHS gene expression data are available at dbGaP under the accession number phs000363.v12.p9.

### Metabolomic profiling

Metabolomics data were available for participants in the PIVUS. Untargeted metabolomic profiling of serum samples was measured in duplicates as described previously<sup>28</sup>. In brief, 1  $\mu$ l of sample was analysed on Acquity UPLC coupled to a Xevo G2 Q-TOFMS (Waters Corporation, Milford, Massachusetts, USA) and raw data was processed using XCMS software<sup>29</sup> for detection, alignment, grouping, and imputation of features. For normalisation of data metabolic feature intensities were log-transformed and an ANOVA-type normalisation applied. Fragmentation spectra were reconstructed from metabolic features with strong correlation and similar retention time and metabolites were identified from spectra. Only annotated metabolites (n=229) were used in analysis in relation to DNA methylation. Raw spectra from mass spectrometry analysis and annotated metabolites intensities are available in Metabolights (<http://www.ebi.ac.uk/metabolights/>) with accession number MTBLS90.

### References

1. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet.* 2013;45:1274-1283
2. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 2013;45:25-33
3. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, Henneman P, Heid IM, Kizer JR, Lyytikainen LP, et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: A multi-ethnic meta-analysis of 45,891 individuals. *PLoS Genet.* 2012;8:e1002607
4. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, Strawbridge RJ, Pers TH, Fischer K, Justice AE, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015;518:187-196
5. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197-206
6. Tsai FJ, Yang CF, Chen CC, Chuang LM, Lu CH, Chang CT, Wang TY, Chen RH, Shiu CF, Liu YM, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in han chinese. *PLoS Genet.* 2010;6:e1000847

7. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The framingham offspring study. *Am J Epidemiol.* 1979;110:281-290
8. Lind L, Fors N, Hall J, Marttala K, Stenborg A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: The prospective investigation of the vasculature in uppsala seniors (pivus) study. *Arterioscler Thromb Vasc Biol.* 2005;25:2368-2375
9. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: Following up the scottish mental surveys of 1932 and 1947. *J Pers Soc Psychol.* 2004;86:130-147
10. Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V, Campbell H, Whalley LJ, Visscher PM, Porteous DJ, et al. The lothian birth cohort 1936: A study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* 2007;7:28
11. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: The lothian birth cohorts of 1921 and 1936. *Int J Epidemiol.* 2012;41:1576-1584
12. Irvin MR, Kabagambe EK, Tiwari HK, Parnell LD, Straka RJ, Tsai M, Ordovas JM, Arnett DK. Apolipoprotein e polymorphisms and postprandial triglyceridemia before and after fenofibrate treatment in the genetics of lipid lowering and diet network (goldn) study. *Circ Cardiovasc Genet.* 2010;3:462-467
13. Joehanes R, Ying S, Huan T, Johnson AD, Raghavachari N, Wang R, Liu P, Woodhouse KA, Sen SK, Tanriverdi K, et al. Gene expression signatures of coronary heart disease. *Arterioscler Thromb Vasc Biol.* 2013;33:1418-1426
14. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing illumina 450k methylation array data. *BMC Genomics.* 2013;14:293
15. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic cpgs in the illumina infinium humanmethylation450 microarray. *Epigenetics.* 2013;8:203-209
16. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, et al. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* 2014;24:1725-1733
17. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: A flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics.* 2014;30:1363-1369
18. Neurauter AA, Bonyhadi M, Lien E, Nokleby L, Ruud E, Camacho S, Aarvak T. Cell isolation and expansion using dynabeads. *Adv Biochem Eng Biotechnol.* 2007;106:41-73
19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics.* 2007;8:118-127
20. Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, Liu C. Removing batch effects in analysis of expression microarray data: An evaluation of six batch adjustment methods. *PLoS One.* 2011;6:e17238
21. Absher DM, Li X, Waite LL, Gibson A, Roberts K, Edberg J, Chatham WW, Kimberly RP. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to cd4+ t-cell populations. *PLoS Genet.* 2013;9:e1003678
22. Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One.* 2009;4:e6098
23. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86

24. Team RC. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015.
25. Zhang X, Johnson AD, Hendricks AE, Hwang SJ, Tanriverdi K, Ganesh SK, Smith NL, Peyser PA, Freedman JE, O'Donnell CJ. Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes. *Hum Mol Genet*. 2014;23:782-795
26. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39:906-913
27. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249-264
28. Ganna A, Salihovic S, Sundstrom J, Broeckling CD, Hedman AK, Magnusson PK, Pedersen NL, Larsson A, Siegbahn A, Zilmer M, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet*. 2014;10:e1004801
29. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. Xcms: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78:779-787